

Quantitative modeling of the function of kinetically driven transcriptional riboswitches

César Parra-Rojas^a, Boris Fürtig^b, Harald Schwalbe^{b,*}, Esteban A. Hernandez-Vargas^{a,*}

^aFrankfurt Institute for Advanced Studies, 60438 Frankfurt am Main, Germany

^bInstitute for Organic Chemistry and Chemical Biology Center for Biomolecular Magnetic Resonance (BMRZ), Johann Wolfgang Goethe-Universität, 60438 Frankfurt am Main, Germany

Abstract

We propose a generalized modeling framework for the kinetic mechanisms of transcriptional riboswitches. The formalism accommodates time-dependent transcription rates and changes of metabolite concentration and permits incorporation of variations in transcription rate depending on transcript length. We derive explicit analytical expressions for the fraction of transcripts that determine repression or activation of gene expression, pause site location and its slowing down of transcription for the case of the (2'dG)-sensing riboswitch from *Mesoplasma florum*. Our modeling challenges the current view on the exclusive importance of metabolite binding to transcripts containing only the aptamer domain. Numerical simulations of transcription proceeding in a continuous manner under time-dependent changes of metabolite concentration further suggest that rapid modulations in concentration result in a reduced dynamic range for riboswitch function regardless of transcription rate, while a combination of slow modulations and small transcription rates ensures a wide range of finely tuneable regulatory outcomes.

1. Introduction

Riboswitches are non-coding RNA sequences that up- or down-regulate gene expression by selectively binding metabolites and subsequent conformational transitions without the intervention of protein factors [1]. They commonly reside in the 5'-untranslated region of bacterial mRNAs. Typically, riboswitches contain two functional domains: an *aptamer domain*, which forms the metabolite-binding pocket, and an *expression platform*, which interacts with the aptamer, assessing the metabolite binding status and modulating gene expression via conformational changes in response to it [1, 2]—see Fig. 1.

Riboswitches can act at the level of transcription or translation. Their regulatory function is achieved mainly by two mechanisms corresponding to translation initiation and rho-independent transcription termination or antitermination. In the case of translation, regulation involves the ribosome binding site (RBS) of the mRNA, which can be either shielded

*To whom correspondence should be addressed.

Emails: schwalbe@nmr.uni-frankfurt.de, vargas@fias.uni-frankfurt.de

from (OFF switch) or accessible to (ON switch) ribosome binding to the mRNA [3] in response to metabolite binding to the upstream aptamer domain. Repression of transcription involves the conditional formation of a terminator stem in response to metabolite binding. This results in premature abortion of transcription. If the terminator is not formed by facilitating of a mutually-exclusive *antiterminator* structure [1, 4], RNA transcription continues. The interaction between the two functional domains, the aptamer domain and the expression platform, relies on the base pairing behavior of generally four complementary nucleotide stretches: P, A, T, and H. In the OFF state, the 5'-aptamer-strand (P) pairs with the aptamer-stabilizing strand (A) and the switching strand (T) forms the terminator helix together with the terminator strand (H). By contrast, in the ON state, A and T or P and T pair and form the antiterminator conformation, and the P/A and H strands are unpaired.

Regulation of transcription by riboswitches is a kinetic phenomenon that cannot be explained by assuming only a few regulatory relevant static riboswitch conformations and the differences of binding affinities at equilibrium to these states. The slow kinetics of RNA refolding after ligand binding play an essential role in the process [5]: if binding does not occur before the terminator stem is synthesized, transcription will proceed. In biological assays, metabolite concentrations 2-3 orders of magnitude higher than the equilibrium dissociation constant K_D are required to effectively regulate expression and, the faster the transcription, the higher the concentration needed in order to trigger termination [6]. Such high affinities of riboswitches for their corresponding metabolites can be exceedingly high, such that an aptamer could potentially detect even a single molecule per cell [4]. This would render gene expression regulation infeasible. Therefore, manipulating the transcription speed by means of *e.g.*, lengthening or shortening the time needed for the formation of the terminator stem conformation [4], or by exploiting pause sites [2], results in tuning the concentration levels at which the riboswitch is able to operate. Incorporation and location of pause sites for transcriptional regulation underlies strict requirements: They have to be placed at specific positions within the transcript, as they have to modulate the time frame for synthesis intervening sequences A-to-T and T-to-H.

The transcription rate not only competes against metabolite binding rates, but also against the refolding rates of RNA itself, and the effects of transcription speed on gene expression can be observed even in total absence of metabolite [7]. These general considerations highlight the importance of the formation of metastable conformations during transcription, and the existence of a well-defined range of transcription speeds at which riboswitch function is possible at all [8]. The simplicity of the regulatory function of riboswitches and the accessibility of all required kinetics and thermodynamics for such regulation at least *in vitro* allows for quantitative modeling of regulation function. Here, we develop a model that integrates all regulatory important factors: transcription speed and its modulation by pause sites, acquisition of RNA structure, metabolite-binding-induced RNA refolding, with their associated timescales, as well as the influence of metabolite concentration, and external variables such as temperature. While the analysis shown here is based on the particular example of an OFF switch, the modeling formalism is applicable to both ON and OFF transcriptional riboswitches in general.

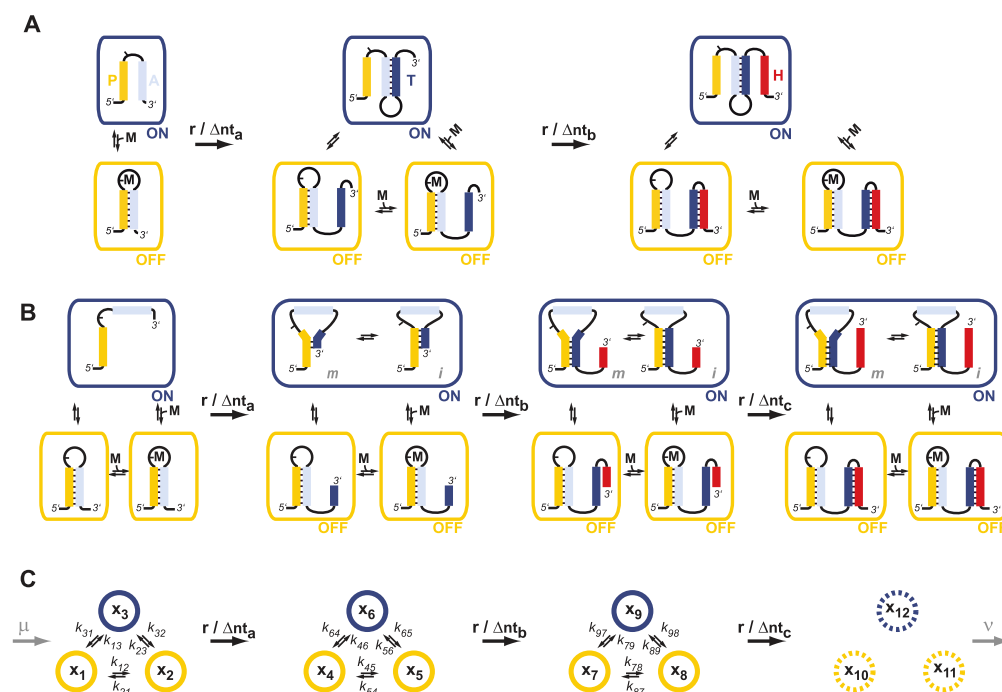


Figure 1: Models of co-transcriptional folding pathways (**A**) Conformational states in context of the transcription progress of a riboswitch with a sequential base-paired stabilized arrangement of the switching sequences P, A, T and H (yellow, grey, blue and red, respectively; arrangement is exemplified by the guanine sensing riboswitch). Metabolite (M, black) binding can occur as soon as the aptamer domain is transcribed and consequently locks A in the PA helix resulting in the population of a single conformation during the transcription process and subsequently in transcription termination. The aptamer domain is synthesized first. As transcription continues, the free mRNA adopts the metastable antiterminator conformation stabilized by AT interaction which refolds to the terminator conformation after the riboswitch is completed. (**B**) Conformational states in context of the transcription progress of a riboswitch with a non-sequential base-pair stabilized arrangement of the switching sequences P, A, T and H (exemplified by the deoxy-guanosine sensing riboswitch). Upon transcription of sequence element T the riboswitch can adopt two fastly interconverting conformations (m and i) in the metabolite free functional ON-state. For both types of riboswitches the functional OFF-state is always characterized by stable interactions between P and A, whereas all functional OFF-states favour non base-paired H sequences. (**C**) The abstraction for the mathematical model, where the m and i ON states are merged into one, so that $N_{on} = 1$ and $N_{off} = 2$. The number of intermediates is $L = 3$, and sink states are introduced at the regulatory decision point, totalling $N = 12$ states for the model. The transcription rate, in nucleotides per second, is denoted by r , while the length of the separation between intermediates, in nucleotides, is denoted by $\Delta nt_{a,b,c}$; at the same time, μ and ν can be thought of as the rates of transcription initiation and transcript degradation, respectively, and are introduced in the model when continuous transcription is considered.

2. Materials and Methods

2.1. Riboswitch folding kinetics

We consider a system of L different transcriptional intermediates, with N_{on} ON states and N_{off} OFF states at each of these steps, the latter including both ligand-free and ligand-bound states—we base the description on an OFF switch, and discuss the ON case below. We will be interested in the fraction of mRNA chains that remain in the ON state during transcription, *i.e.*, the fraction ON among the full-length transcripts. In practice, this means that we need to allow for complete conversion of the shorter transcription intermediates. We introduce *sink* states at the end of a single round of transcription, that will accumulate the transcripts in ON and OFF states after the regulatory decision point has been reached and transitions between states are no longer possible or no longer biologically relevant in terms of gene expression. In mathematical terms, the quantity of interest corresponds to the combined steady-state values of the ON sinks. The total number of states in the system will correspond to $N = N_L(L + 1)$, where $N_L = N_{on} + N_{off}$ is the total number of states per transcript length. The fraction of transcripts occupying state i at a given time t will be denoted by $x_i(t)$, so that the occupation levels in the model are characterized by a vector $\mathbf{x}(t) = (x_1(t), \dots, x_N(t))$. An illustration of such a system with $L = 3$, $N_{on} = 1$, and $N_{off} = 2$ —thus, $N = 12$ —is shown in Fig. 1. An important distinction must be made between this formulation and our previous works [9, 7, 8], which do not consider a final set of states and therefore were simulated up to a *stopping time* given by the mean elongation time from the shortest to the longest transcripts.

The general form of the kinetics is given by

$$\dot{\mathbf{x}} = \mathbf{A} \cdot \mathbf{x}(t), \quad (1)$$

where \mathbf{A} is an $N \times N$ matrix with entries

$$\begin{aligned} A_{ij} &= k_{ji}, & i \neq j \\ A_{ii} &= -\sum_j k_{ij}, \end{aligned} \quad (2)$$

and k_{ij} represents the folding rate from state i to state j . Here, we assume that the folding rates between conformations have been either measured experimentally or predicted *in silico*—see, *e.g.*, Flamm *et al.* (2002) [10]. The number of relevant intermediates, L , and the distance in nucleotides between any given pair of them, is chosen *a posteriori* so that the k_{ij} are all constant. This means that, *e.g.*, in Fig. 1 the folding rates were measured and found to remain constant or approximately constant for Δn_{t_a} nucleotides, then changing and remaining the same for another Δn_{t_b} nucleotides, and so on.

We will adopt the convention that states are ordered from short to long, from OFF to ON so that, for a system of size N , we will have

$$\begin{aligned} (x_1, \dots, x_N) &= (x_{off,1}^{(1)}, x_{off,2}^{(1)}, \dots, x_{off,N_{off}}^{(1)}, x_{on,1}^{(1)}, \dots, x_{on,N_{on}}^{(1)}, \dots \\ &\quad \dots, x_{off,1}^{(L)}, \dots, x_{on,N_{on}}^{(L)}, x_{off,1}^{(s)}, \dots, x_{on,N_{on}}^{(s)}), \end{aligned} \quad (3)$$

where s stands for ‘sink’. We arrange the state vector in the mathematical description so that, per transcript intermediate, all ligand-free states come before ligand-bound states. From this, it follows that the quantity of interest will be the last N_{om} elements of the state vector when $t \rightarrow \infty$. If we decompose the label of a certain state i as $i = p_i N_L + q_i$, with $p_i = 0, 1, \dots, L$ and $q_i = 1, 2, \dots, N_L$, then the following properties are satisfied:

1. All transitions rates k_{ij} for which $p_i > p_j$ are equal to zero; that is, transitions from longer to shorter transcripts are forbidden.
2. Transitions for which $p_i < p_j$ and $q_i \neq q_j$ are also equal to zero, since the lengthening of the transcripts preserves their state. In other words, the lengthening of a transcript in, *e.g.*, state $x_{off,1}^{(1)}$ can only take it to state $x_{off,1}^{(2)}$ but not to, say, $x_{off,2}^{(2)}$.
3. Transitions for which $q_i = q_j$ but $p_j - p_i > 1$ are also equal to zero, due to the fact that the lengthening is sequential.
4. By construction, the transition rates between two given transcript lengths are all equal and independent of the state, so that $k_{ij} = k_{lm}$ for all cases when $p_i = p_l$ and $p_j = p_m$. Furthermore, these rates are uniquely determined by the transcription rate, which we denote by r ; r has units of $\text{nt} \cdot \text{s}^{-1}$, and is defined as $r = 1 [\text{nt}] \cdot K [\text{s}^{-1}]$, where K represents the rate of nucleotide addition by the polymerase, in s^{-1} .

2.1.1. Analytical solutions

The model can be explored numerically, both for constant and time-varying transcription rates. However, analytical solutions can be found if the transcription rate is a constant and the total number of states is sufficiently small. The general solution reads

$$\mathbf{x}(t) = \sum_{i=1}^N c_i \mathbf{v}_i e^{\lambda_i t},$$

where λ_i and \mathbf{v}_i are, respectively, the i -th eigenvalue and the i -th right eigenvector of \mathbf{A} , while the c_i , $i = 1, \dots, N$ are coefficients that are uniquely determined by the initial condition $\mathbf{x}_0 = \mathbf{x}(t = 0)$. Using the expression above, these can be found from

$$\mathbf{x}_0 = \sum_{i=1}^N c_i \mathbf{v}_i. \quad (4)$$

We note that, by construction, the system with the sink states will have N_L eigenvalues equal to zero with eigenvectors being the unit vectors corresponding to the sinks— $v_{i,j} = \delta_{ij}$ for $i = N_L L + 1, \dots, N$ —and with all other eigenvalues $\lambda_i < 0$. Thus, its steady-state solution, which we denote by a vector $\bar{\mathbf{x}}$, will be given by

$$\bar{x}_i = c_i, \quad i = N_L L + 1, \dots, N, \quad (5)$$

with $\bar{x}_i = 0$ otherwise.

Let us consider the (2’dG)-sensing riboswitch from *Mesoplasma florum* as an illustration. Previously, we have determined experimentally the relevant transcriptional intermediates

and the folding rates between conformations for this riboswitch [8]. In principle, this system consists of two ON states per transcriptional intermediate, with $L = 2$ transcriptional intermediates; however, if we are interested in the total fraction of transcripts allowing gene expression, the system size effectively reduces from twelve to nine dimensions. This is analogous to Fig. 1B, showing two ON states merged into one for a system with $L = 3$, reduced from 16 to 12 states. As in the original model, let us assume that there is a binding window of length $\Delta nt_a \equiv 1/a$ that has already been traversed, with a resulting fraction b_0 of bound transcripts. That is, we consider only the states x_4 to x_9 of the nine-dimensional representation, with an initial condition given by: $x_4 = 1 - b_0$, $x_5 = b_0$, $x_i = 0$ for $i = 6, \dots, 9$. This corresponds to a single round of transcription starting from its shortest length— $x_i = 0$ for $i = 7, 8, 9$ —and in a fully OFF configuration— $x_6 = 0$ —with a fraction b_0 of bound transcripts. Folding is only possible from OFF to ON states—the reverse rates are negligible [8]—so that we set $k_{45} = k_{54} = k_{64} = k_{65} = 0$, and an explicit solution can be found for the fraction ON at full length:

$$\bar{x}_9 = (1 - b_0) \frac{k_{on}^{(f)}}{k_{on}^{(f)} + br} + b_0 \frac{k_{on}^{(b)}}{k_{on}^{(b)} + br}, \quad (6)$$

where $\Delta nt_b \equiv 1/b = 24 \text{ nt}$ is the length of the folding window from nucleotide 113 to nucleotide 137, $k_{46} = k_{on}^{(f)}$ is the folding rate from the ligand-free OFF state to the ON state, and $k_{56} = k_{on}^{(b)}$ is the folding rate from the ligand-bound OFF state to the ON state. We may compare this result to the original [8], given by

$$x_3(T) = 1 - e^{-k_{on}^{(f)}T} + b_0 \left(e^{-k_{on}^{(f)}T} - e^{-k_{on}^{(b)}T} \right), \quad (7)$$

with $T = (br)^{-1} \text{ s}$, corresponding to the mean time it takes the transcription to proceed by 24nt.

2.2. Pause sites

We start by noting that changes in the stability of the different conformations enter the model through the addition or removal of intermediates of a given length. In other words, the intermediates are defined by windows of approximately constant folding rates so that, by construction, all rates k_{ij} appearing in \mathbf{A} —Eq. (2)—and relating to transitions between conformations are constant, while elongation transitions are directly defined through the transcription rate r .

For the model systems described here, we have previously shown that the transcription rate experiences changes at specific transcript lengths. In this case we can exploit the same strategy as for changes in folding rates: we can insert a window of reduced transcription rate as another transcriptional intermediate in the system. In other words, in order to model a pause site located, *e.g.*, in the last quarter of a window of length Δnt , we can insert an intermediate that divides the window into one sub-window of length $3\Delta nt/4$ and another of $\Delta nt/4$, with the latter being transcribed at a (much) lower rate. In general, changes in transcription rate at specific locations can be accommodated by further subdivisions.

2.3. Metabolite concentration changes

Another factor that influences the dynamics of the system is the metabolite concentration, which will have an impact on the fraction of bound transcripts at the end of the binding window, and that we have so far not considered. Since the metabolite does not play a role in the fate of the transcripts once they are past the binding window, in order to explore the consequences of varying its concentration over slower timescales we need to modify the model to account for continuous transcription. We can picture this as a system that is producing transcripts at a rate μ , that we may identify with the rate of *transcription initiation*; at the same time, we assume that full-length transcripts leave the system at a rate ν , signifying *transcript degradation*—see Fig. 1. We make a simplification by treating the metabolite concentration M as an external reservoir we can modify arbitrarily; that is, it does not depend on the dynamics of the riboswitch. Additionally, for the sake of simplicity, we take the degradation rate to be the same as the initiation rate, $\nu = \mu$. Under this condition, the total sum of full-length transcripts reaches unity at equilibrium, and we will be interested in the fraction of full-length transcripts in ON state once this equilibrium has been reached.

We initialize the system in absence of metabolite, and start monotonically increasing its concentration at a certain time t_0 up to a time t_1 , such that if $t_1 \gg t_0$ the concentration reaches a saturation value which we choose to be $M = 1$. At time t_1 , we begin to decrease the concentration until no metabolite is present in the system. We use

$$M(t) = \frac{(t - t_0)^2}{1 + (t - t_0)^2} \Theta(t - t_0) \Theta(t_1 - t) + \left[\frac{(t - t_0)^2}{1 + (t - t_0)^2} - \frac{(t - t_1)^2}{1 + (t - t_1)^2} \right] \Theta(t - t_1), \quad (8)$$

where $\Theta(\cdot)$ corresponds to the Heaviside theta function.

3. Results

3.1. Full conversion of shorter transcripts in the (2'dG)-sensing riboswitch from *Mesoplasma florum* results in a lower fraction ON

The analytical solution for the fraction ON at a length of 137 nucleotides in the (2'dG)-sensing riboswitch from *Mesoplasma florum* is shown in Eqs. 6 and 7 for the formulations with and without sink states, respectively. Comparing both results, we observe a generally lower fraction ON in the case with sinks—see Fig. 2. We can explain this discrepancy by noting that, in the original formulation [8], all transcripts at nucleotide 113 have a life-time that is exactly equal to the mean time to elongation by 24nt, T . By contrast, in the case with sinks, the life-times are exponentially distributed with the same mean. As a result, the majority of the transcripts will be elongated before a time T has passed, and have therefore a lower chance of folding towards the ON state. This is a subtle but important conceptual difference between this and our previous works [9, 7, 8].

We have started from a specified fraction of ligand-bound transcripts for the purpose of the comparison. However, the full result must take into account the fact that this fraction

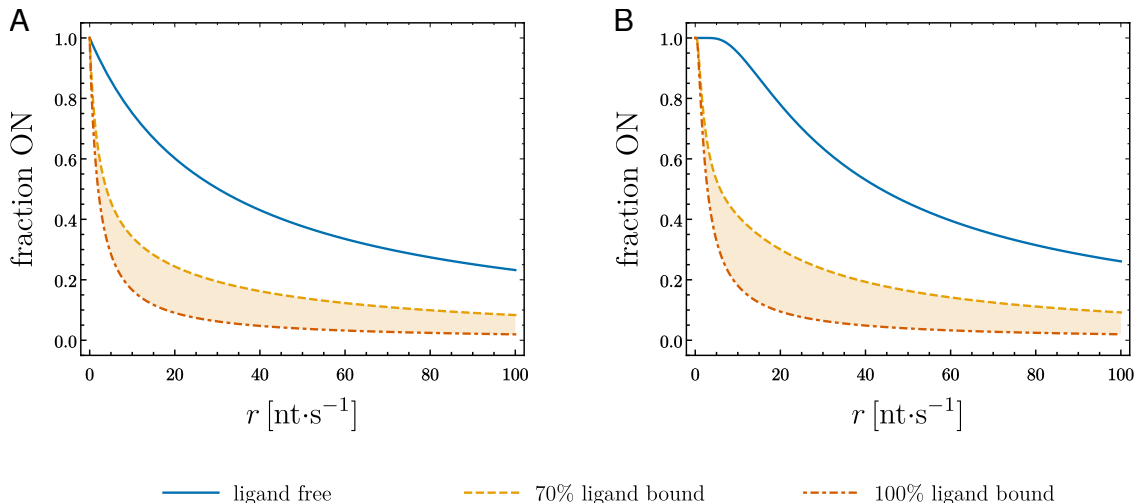


Figure 2: Fraction of transcripts in the ON conformation at a length of 137 nucleotides in a single round of transcription for the (2'dG)-sensing riboswitch from *Mesoplasma florum*, as a function of constant transcription rate r and a fixed fraction of initially bound transcripts. **(A)** With full conversion of shorter transcripts through the addition of sink states. **(B)** In the original formulation without the sink states, from Helmling *et al.* (2018) [8].

is the result of binding events between the first and second transcriptional intermediates—see Fig. 1. Therefore, we include the dynamics during the binding window, with allowed transitions only from the unbound to the bound OFF state— $k_{13} = k_{21} = k_{23} = k_{31} = k_{32} = 0$ in Fig. 1. We take $k_{12} = k_b M$, where k_b is the metabolite binding rate and we have assumed that the ligand concentration, M , is large enough so that binding events have a negligible effect on it. In this case, $1/a = 35$ nt, and $1/b = 24$ nt, corresponding to the length of the binding and folding windows, respectively. The full solution for \bar{x}_9 , corresponding to the fraction ON for a transcript length of 137 nucleotides in a single round of transcription— $\mu = \nu = 0$ —takes the form

$$\bar{x}_9 = \frac{1}{k_b M + a r} \left(\frac{a r k_{on}^{(f)}}{k_{on}^{(f)} + b r} + \frac{k_b M k_{on}^{(b)}}{k_{on}^{(b)} + b r} \right). \quad (9)$$

The full landscape of this solution is plotted in Fig. 3A as a function of a constant transcription rate r and the product $k_b M$. It is straightforward to see that the limits $k_b M \rightarrow 0$ and $k_b M \rightarrow \infty$ result, respectively, in no ligand binding and 100% ligand binding, corresponding to the cases $b_0 = 0$ and $b_0 = 1$ from Eq. 6. For vanishingly slow transition rates, the resulting fraction ON tends to unity regardless of the ligand concentration and binding rate.

Something the results show is that solutions can be found by iteration. Suppose that we focus only on the binding window of length $1/a$, and “stop” the dynamics after it is traversed. If we identify b_0 as the solution for x_5 when $t \rightarrow \infty$ if we forbid all transitions between x_4 , x_5 and x_6 , we obtain

$$b_0 = \frac{k_b M}{k_b M + a r}, \quad (10)$$

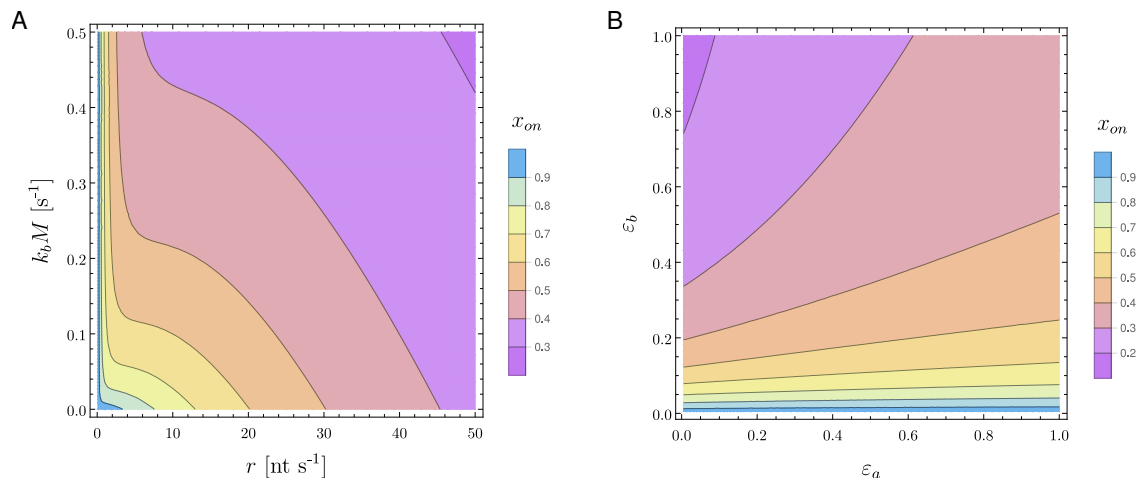


Figure 3: Contour lines for the fraction ON at a length of 137 nucleotides in a single round of transcription for the (2’dG)-sensing riboswitch from *Mesoplasma florum*—same as Fig. 1B, but with $L = 2$, $N = 9$ —with folding rates taken from Helmling *et al.* (2018) [8]. **(A)** As a function of constant transcription rate r and the product of the ligand binding rate and the ligand concentration $k_b M$, obtained from Eq. 9. **(B)** With binding (length $1/a$) and folding (length $1/b$) windows divided into two sections traversed at different transcription rates, $r_{1,2}^{(a,b)}$. Result obtained from Eq. 11 when the rates follow $\bar{r}_2^{(a)} = \varepsilon_a \bar{r}_1^{(a)}$, and $\bar{r}_2^{(b)} = \varepsilon_b \bar{r}_1^{(b)}$, and are expressed as a function as pause site location and its amount of slowing down: respectively, either at the binding ($\varepsilon_a < 1$) or folding ($\varepsilon_b < 1$) window, and the exact magnitude of $\varepsilon_{a,b}$. Base parameters are $r_1^{(a)} = r_1^{(b)} = 10 \text{ nt s}^{-1}$, $w_1^{(a)} = 18 \text{ nt}$, $w_1^{(b)} = 12 \text{ nt}$, and $k_b M = 10/18 \text{ s}^{-1}$.

and Eq. 9 reduces to Eq. 6. In other words, a full analytical solution for intermediate states at a given length may be found recursively from the solution to the fraction of transcripts that eventually pass through the states at the previous length.

3.2. Pause sites located after the binding window increase the switching efficiency

As an illustration, both the binding and folding windows of the system from the previous section are divided into two sub-windows, the lengths of which we denote by w_1 and w_2 ; that is, $\Delta \text{nt}_a = 1/a = w_1^{(a)} + w_2^{(a)}$, and $\Delta \text{nt}_b = 1/b = w_1^{(b)} + w_2^{(b)}$. We assume that, in each case, the sub-window w_1 is transcribed at rate r_1 and the sub-window w_2 is transcribed at rate r_2 . We are interested in the effects of the location of pausing events, which will enter as a difference between the rates $r_{1,2}$ in the binding window, the folding window, or both. We are also interested in the intensity of the slowing down in transcription, which will correspond to how much lower is one of the rates $r_{1,2}$ with respect to the other.

This modified system has a total of 15 states, and the steady-state solution for the final

fraction ON in a single round of transcription is given by

$$\begin{aligned} \bar{x}_{15} = & \frac{1}{\left(\bar{r}_1^{(a)} + k_b M\right) \left(\bar{r}_2^{(a)} + k_b M\right)} \\ & \times \left\{ \bar{r}_1^{(a)} \bar{r}_2^{(a)} \left[1 - \frac{\bar{r}_1^{(b)} \bar{r}_2^{(b)}}{\left(\bar{r}_1^{(b)} + k_{on}^{(f)}\right) \left(\bar{r}_2^{(b)} + k_{on}^{(f)}\right)} \right] \right. \\ & \left. + k_b M k_{on}^{(b)} \frac{\left[\left(\bar{r}_1^{(a)} + k_b M\right) + \bar{r}_2^{(a)}\right] \left[\left(\bar{r}_1^{(b)} + k_{on}^{(b)}\right) + \bar{r}_2^{(b)}\right]}{\left(\bar{r}_1^{(b)} + k_{on}^{(b)}\right) \left(\bar{r}_2^{(b)} + k_{on}^{(b)}\right)} \right\}, \end{aligned} \quad (11)$$

where $\bar{r}_{1,2}^{(a,b)} \equiv r_{1,2}^{(a,b)}/w_{1,2}^{(a,b)}$. This solution reduces to Eq. 9 when all the transcription rates are equal, one of the binding windows takes the full length $1/a$, and one of the folding windows takes the full length $1/b$.

Since a more pronounced slowing down of transcription has the same effect as lengthening the pausing window—note that the terms involving the transcription rates, $r_{1,2}^{(a,b)}$, appear only in combination with those corresponding to window length, $w_{1,2}^{(a,b)}$ —we can explore the effects of pausing strength and extension at the same time. Let us suppose that $\bar{r}_2^{(a)} = \varepsilon_a \bar{r}_1^{(a)}$, and $\bar{r}_2^{(b)} = \varepsilon_b \bar{r}_1^{(b)}$ —that is, the effective transcription rate at one portion of the binding/folding window is a well-defined fraction of the one at the other portion. Fig. 3B shows the resulting fraction ON from Eq. 11 as a function of this fraction. We see from this that stronger pause events localized in the binding window tend to result in a lower fraction ON, due to the transcripts having more time to bind the metabolite. Stronger pauses during the folding window, in turn, have the opposite effect, since the transcripts have more time to refold from the OFF to the ON state. In agreement with the results depicted in Fig. 3A, a sufficiently intense reduction in transcription rate during folding results in the fraction ON being close to unity: $\bar{x}_{15} \rightarrow 1$ in Eq. 11 for $\varepsilon_b \rightarrow 0$, regardless of the value of ε_a .

We have explored the simple case of a pause extending from the beginning of the window until a certain point (resp. from a certain point until the end). The more realistic scenario of a short localized pause can be studied using the recursive property of the solutions that we have discussed above. However, we expect the qualitative insights gleaned from Eq. 11 to be valid in this case as well.

3.3. Fast variations in metabolite concentration reduce the dynamic range

As before, we take the folding rates from our previous work [8]. We further choose $k_b = 1$ so that, for a constant metabolite concentration $M = 1$ and a constant transcription speed $r = 35 \text{ nt s}^{-1}$, the fraction of bound transcripts at nucleotide 113 is $b_0 = 0.5$ if initiation and degradation are ignored—see Eq. 10. As an example, we take the metabolite concentration to satisfy Eq. 8. Fig. 4 shows $M(t)$, along with the resulting dynamics for the fraction ON at full-length for a choice of t_0 and different values of t_1 and transcription rate. In this illustration we see that, the larger the transcription rate, the faster the system tends to react to changes in metabolite concentration. However, at the same time, the folding window is traversed too quickly for conformational changes to occur to a great extent even for vanishing

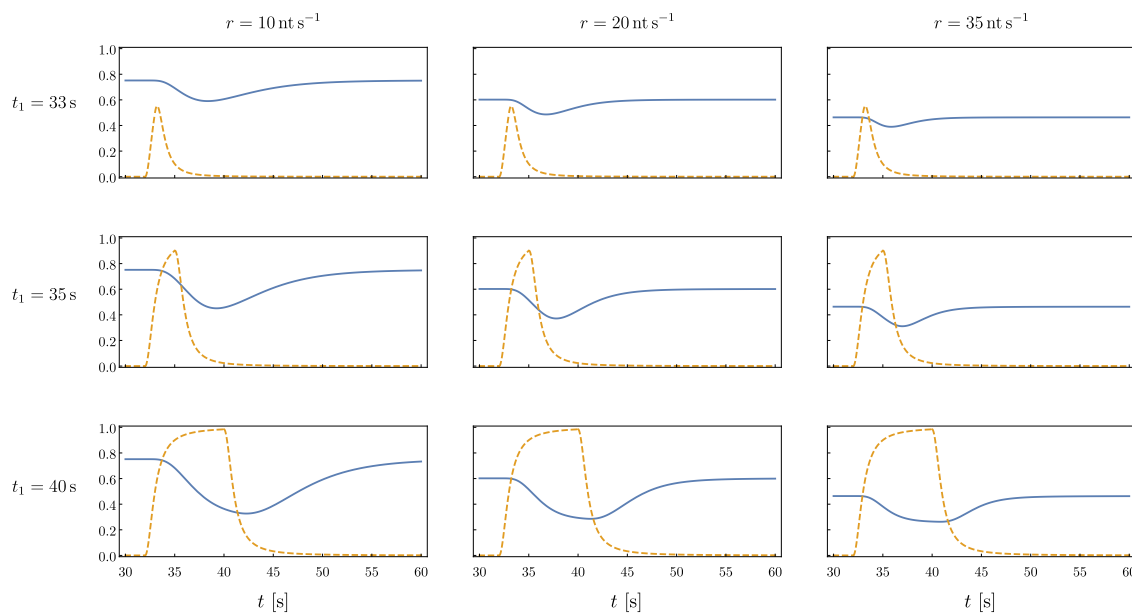


Figure 4: **Orange, dashed line:** time-varying ligand concentration from Eq. 8 for $t_0 = 32$ s; top: $t_1 = 33$ s; middle: $t_1 = 35$ s; bottom: $t_1 = 40$ s. **Blue, solid line:** resulting fraction ON at a length of 137 nucleotides for the (2'dG)-sensing riboswitch from *Mesoplasma florum* under continuous transcription at a constant rate; left: $r = 10$ nt s⁻¹; center: $r = 20$ nt s⁻¹; right: $r = 35$ nt s⁻¹. An equilibration time of 30 s has been used. Parameter values: $k_b = \mu = \nu = 1$. Refolding rates from Helmling *et al.* (2018) [8].

M , so that the dynamic range of the output is reduced as we increase the transcription rate. On the other hand, smaller transcription rates result in a slower reaction time but permit increased refolding. In this case, slower modulations in M allow the system to equilibrate at a larger fraction of bound transcripts at the end of the binding window. The combination of both larger concentration equilibration times and slower transcription rates results in a wider dynamic range for riboswitch function.

4. Discussion

Naturally-occurring riboswitches are commonly found in bacteria, where they take part in the regulation of the expression of a number of genes which are essential for survival [11], as well as directly influencing motility and pathogenesis [12]. This, coupled with the selectivity with which they bind their corresponding metabolites, makes them attractive drug targets for the design of novel antibiotics [13, 11]. Engineered riboswitches, in turn, have been successfully used to control the replication and infectivity of both RNA and DNA viruses [14, 15]. This highlights a variety of potential therapeutic applications of riboswitch regulation, and the need for a deep qualitative and quantitative understanding of the mechanisms involved.

Mathematical approaches to riboswitch kinetics are scarce in the literature: examples include Santillán & Mackey (2005) [16], who constructed a dynamical model for the B₁₂-sensing *btuB* riboswitch from *E. coli* and analyzed transient and steady states in terms of B₁₂ levels subject to both transcriptional and translational regulation; Beisel & Smolke (2009) [17],

who qualitatively outline the general performance properties, in terms of protein levels, of ON and OFF riboswitches under different regulatory mechanisms, and their implications for the tuning of natural and synthetic riboswitches; and Hofacker *et al.* (2010) [18], who describe the dynamics of RNA co-transcriptional folding in a step-by-step formulation through a set of transcription stopping times.

Here, we have presented a generalized modeling framework for the kinetics of transcriptional riboswitches, based on a group of relevant transcriptional intermediates defined by windows of approximately constant (re)folding rates, with a number of states at each step defined by the different available structural conformations. We assume that these windows are determined experimentally, and the relevant intermediates are subsequently chosen based on them. We introduce a final set of states, with no folding transitions between them, so that the output of the model corresponds to fraction of transcripts that eventually exit a transcription round in a given conformation. The formulation of the model is general and simple enough to accommodate either ON or OFF switches, and we note that, while alternatives for a full numerical characterization of co-transcriptional folding are available [19], our aim is to provide a framework to readily try out different configurations in a system with known, few relevant states, which permits both qualitative and quantitative analyses of the co-transcriptional interplay between transcription and folding rates, as well as metabolite concentration and binding rate. The sequential character of the model allows for a closed-form solution to be obtained at each intermediate step for sufficiently simple conformation configurations, *e.g.*, through the use of computer algebra [20], while the full solution can be subsequently pieced together from its component parts in a recursive manner. Furthermore, the existence of closed-form solutions facilitate the direct qualitative and quantitative exploration of the behavior of *classes* of riboswitches sharing their main structure and differing only on their distance between intermediates or their folding rates.

As an illustration, we have derived a number of analytical expressions for the fraction of transcripts in ON state at the regulatory decision point for a simple system, and applied them to the explicit case of the (2'dG)-sensing riboswitch from *Mesoplasma florum*; the model is able to reproduce previous results when the condition of short-transcript conversion is relaxed [8]. Due to the recursive character of the solutions, we have also shown how changes in conformation stability and position-dependent changes in transcription rate can be incorporated in the model through additional intermediates without rendering the system intractable. With this, we derived an analytical expression for the landscape of possible outcomes representing the effects of pause site strength and location in the system. For the case of the (2'dG)-sensing riboswitch from *Mesoplasma florum*, pausing at latter stages in transcription has a much stronger effect on the outcome than pausing during the binding window: while metabolite concentration and binding rate do play an important role in the overall dynamics, a sufficiently small transcription rate or sufficiently pronounced pausing event after binding is no longer possible gives the system enough time to almost entirely refold to the ON state regardless of the fraction of bound transcripts.

From numerical simulations of continuous transcription in presence of external variations in metabolite concentration, we have observed that the relevant timescale for riboswitch

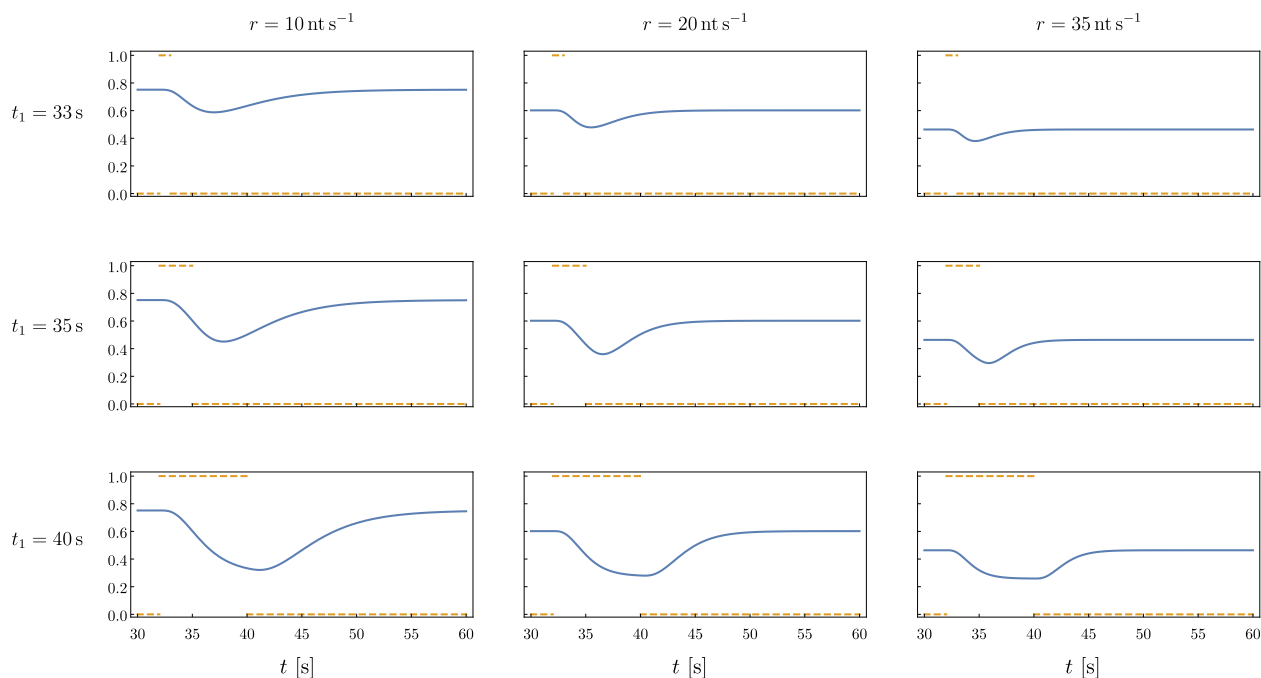


Figure 5: **Orange, dashed line:** the ligand concentration follows a step function, being set to $M = 1$ from a time $t_0 = 32$ s until a time t_1 , and to $M = 0$ otherwise; top: $t_1 = 33$ s; middle: $t_1 = 35$ s; bottom: $t_1 = 40$ s. **Blue, solid line:** resulting fraction ON at a length of 137 nucleotides for the (2'dG)-sensing riboswitch from *Mesoplasma florum* under continuous transcription at a constant rate; left: $r = 10 \text{ nt s}^{-1}$; center: $r = 20 \text{ nt s}^{-1}$; right: $r = 35 \text{ nt s}^{-1}$. An equilibration time of 30 s has been used. The resulting behavior is equivalent to that of Fig. 4. Parameter values: $k_b = \mu = \nu = 1$. Refolding rates from Helmling *et al.* (2018) [8].

function corresponds to the mean time it takes for the binding window to be traversed, and its relationship to the timescale of changes in metabolite concentration. For slow changes in concentration the system equilibrates at large bound fractions, and the dynamic range of the final fraction ON depends entirely on the transcription rate. This agrees with our intuition and analytical results since, even in total absence of metabolite, if transcription proceeds too quickly the transcripts do not have sufficient time to refold from the OFF to the ON state, so that dynamic range is reduced at a small fraction ON—bottom-right panel of Fig. 4. At the same time, for sufficiently low transcription rates, permitting a large fraction ON in absence of metabolite, fast modulations in concentration result in a reduced dynamic range at a large fraction ON—top-left panel of Fig. 4. A combination of both slow changes in metabolite concentration and low transcription rates, in turn, yields a wide dynamic range—bottom-left panel of Fig. 4. This is not affected by the maximum concentration attained by the metabolite at different values of the stopping time t_1 , and using a step function yields the same behavior—see Fig. 5.

While our illustrations have been based on the case of an OFF riboswitch, for which metabolite binding stabilizes the OFF state—thus drastically reducing the folding rate towards the ON state—the analytical expressions we have derived are defined in terms of

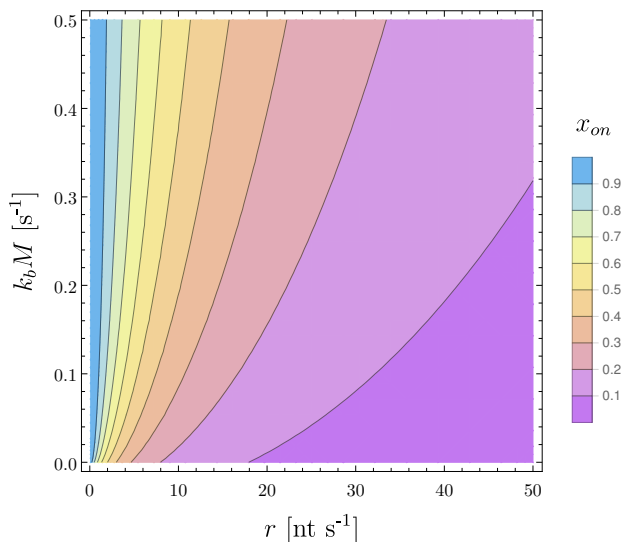


Figure 6: Contour lines for the fraction ON at a length of 137 nucleotides in a single round of transcription for the (2'dG)-sensing riboswitch from *Mesoplasma florum*, as a function of constant transcription rate r and the product of the ligand binding rate and the ligand concentration $k_b M$, when the folding rates from the ligand-free and ligand-bound OFF states to the ON state have been swapped, thus resulting in an ON switch. Parameter values from Helmling *et al.* (2018) [8].

arbitrary values for the folding rates $k_{on}^{(f;b)}$ so that, without modifying the available conformations and the allowed transitions between them in the *Mesoplasma florum* example, we can accommodate a simple ON switch that destabilizes the OFF state upon metabolite binding by simply making $k_{on}^{(f)} \ll k_{on}^{(b)}$. Figure 6 shows the same landscape as Fig. 3A for the case of an ON switch when the values of $k_{on}^{(f)}$ and $k_{on}^{(b)}$ are swapped; Fig. 7, in turn, depicts both cases, OFF and ON, as a function of the metabolite-induced stabilization/destabilization strength—determined by the ratio between $k_{on}^{(f)}$ and $k_{on}^{(b)}$ —for a fixed transcription rate. The different effects of increased metabolite binding—as a result of higher concentration and/or binding rate—for the two cases can be clearly appreciated. In general, the more stable one configuration is with respect to the other, the larger is the dynamic range of the outcome.

We have explored the dynamics of the riboswitch under a time-dependent metabolite concentration that is manipulated externally, consisting of a single round of growth, equilibration and decay. While other slower-than-transcription effects—such as temperature dependence of transcription or folding rates [21, 22, 23]—can be directly introduced in the framework of continuous transcription without additional steps, the model we have defined may be subsequently coupled to an equation describing the dynamics of metabolite concentration itself as a function of the fraction of transcripts allowing gene expression, in a way akin to that in Santillán & Mackey (2005) [16]. Such a coupling would make possible a full exploration of the feedback mechanisms governing the regulatory behavior of transcriptional riboswitches.

We have also worked under the assumption that the elongation times between transcriptional intermediates are all exponentially distributed, so that the dynamics can be well described by a low-dimensional system of ordinary differential equations (ODEs). The problem

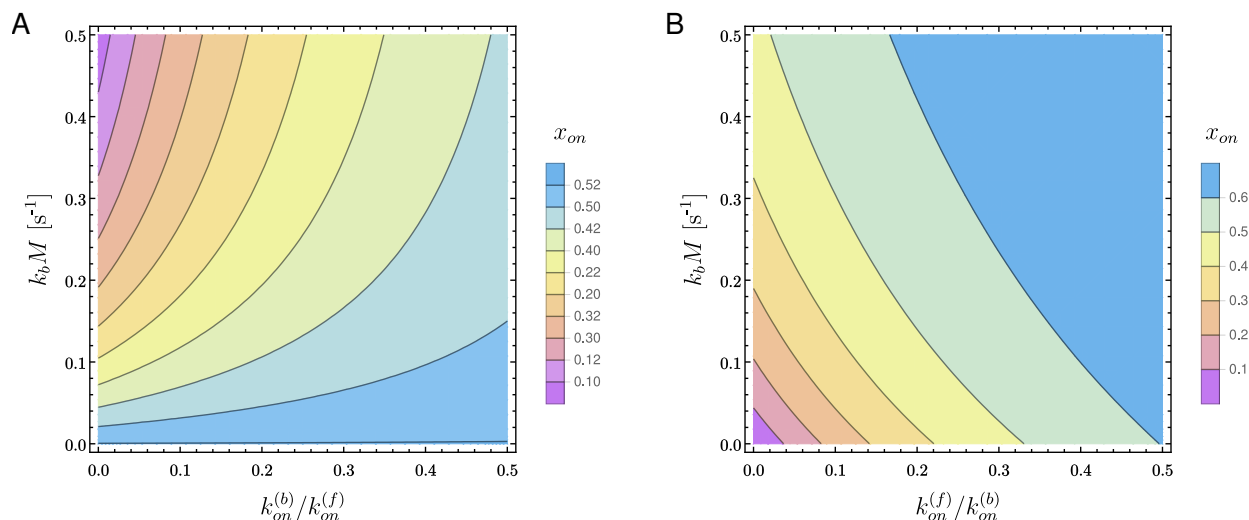


Figure 7: Contour lines for the fraction ON at a length of 137 nucleotides in a single round of transcription for the (2′dG)-sensing riboswitch from *Mesoplasma florum*, as a function of the product of the ligand binding rate and the ligand concentration $k_b M$ and the metabolite-induced stabilization/destabilization of the OFF conformation—determined by the ratio between $k_{on}^{(f)}$ and $k_{on}^{(b)}$ —for a fixed transcription rate $r = 10$ nt s⁻¹. **(A)** Metabolite binding stabilizes the OFF conformation (OFF switch). **(B)** Metabolite binding destabilizes the OFF conformation (ON switch). In both cases, the highest folding rate is set to 1.26 s⁻¹.

may be more complex to model: if we consider single nucleotide addition by the polymerase to be a stochastic process with an exponentially-distributed waiting time, the elongation time from one transcriptional intermediate to another follows a Gamma distribution with a mean equal to the sum of all the individual waiting times [24]; in both cases, transcription proceeds with the same “speed”. We can observe this directly from Eq. 11: even if we choose all transcription rates to be equal, the subdivision of the nucleotide windows into $n > 1$ different sections results in a different form for the fraction ON, compared to that of Eq. 9. For sufficiently large numbers of nucleotides separating all the different transcriptional intermediates, the transitions may be well approximated by a fixed waiting time, as in Hofacker *et al.* (2010) [18]. Within the ODE framework presented here, a full solution may still be obtained recursively from Eq. 9 for the more realistic case comprising windows of 1nt each, traversed in exponentially-distributed times.

5. Conclusion

Riboswitches are cis-acting regulatory mRNA elements in bacteria, that modulate the expression of their associated genes in response to a cognate metabolite. They operate either on the level of translation or transcription. Transcriptional riboswitches have to fold into functional structures as they are being synthesized. Here, we developed a mathematical treatment of the kinetic mechanisms of co-transcriptional riboswitch folding. It permits the evaluation of different transcriptional speeds and transcriptional pause sites on population of co-transcriptional folding intermediates and subsequently the functional outcome. The math-

emational treatment of the kinetic mechanisms further allows the incorporation of metabolite changes in modelling of the folding pathways. Ultimately, our modelling framework will allow to describe riboswitch based-regulation of transcription at nucleotide resolution.

References

- [1] M. Mandal, R. R. Breaker, Gene regulation by riboswitches, *Nature Reviews Molecular Cell Biology* 5 (2004) 451.
- [2] A. D. Garst, A. L. Edwards, R. T. Batey, Riboswitches: structures and mechanisms, *Cold Spring Harbor Perspectives in Biology* 3 (2011) a003533.
- [3] R. R. Breaker, Riboswitches and translation control, *Cold Spring Harbor Perspectives in Biology* 10 (2018) a032797.
- [4] R. R. Breaker, Riboswitches and the RNA World, *Cold Spring Harbor Perspectives in Biology* 4 (2012) a003566.
- [5] J. K. Wickiser, M. T. Cheah, R. R. Breaker, D. M. Crothers, The kinetics of ligand binding by an adenine-sensing riboswitch, *Biochemistry* 44 (2005) 13404–13414.
- [6] J. K. Wickiser, W. C. Winkler, R. R. Breaker, D. M. Crothers, The speed of RNA transcription and metabolite binding kinetics operate an fmn riboswitch, *Molecular Cell* 18 (2005) 49–60.
- [7] C. Helmling, A. Wacker, M. T. Wolfinger, I. L. Hofacker, M. Hengesbach, B. Fürtig, H. Schwalbe, NMR structural profiling of transcriptional intermediates reveals riboswitch regulation by metastable RNA conformations, *Journal of the American Chemical Society* 139 (2017) 2647–2656.
- [8] C. Helmling, D.-P. Klötzner, F. Sochor, R. A. Mooney, A. Wacker, R. Landick, B. Fürtig, A. Heckel, H. Schwalbe, Life times of metastable states guide regulatory signaling in transcriptional riboswitches, *Nature Communications* 9 (2018) 944.
- [9] H. Steinert, F. Sochor, A. Wacker, J. Buck, C. Helmling, F. Hiller, S. Keyhani, J. Noeske, S. Grimm, M. M. Rudolph, et al., Pausing guides RNA folding to populate transiently stable RNA structures for riboswitch-based transcription regulation, *eLife* 6 (2017) e21297.
- [10] C. Flamm, I. L. Hofacker, P. F. Stadler, M. T. Wolfinger, Barrier trees of degenerate landscapes, *Zeitschrift für physikalische chemie* 216 (2002) 155.
- [11] K. F. Blount, R. R. Breaker, Riboswitches as antibacterial drug targets, *Nature Biotechnology* 24 (2006) 1558.

- [12] R. Tamayo, Cyclic diguanylate riboswitches control bacterial pathogenesis mechanisms, *PLOS Pathogens* 15 (2019) 1–7. URL: <https://doi.org/10.1371/journal.ppat.1007529>. doi:10.1371/journal.ppat.1007529.
- [13] N. Sudarsan, S. Cohen-Chalamish, S. Nakamura, G. M. Emilsson, R. R. Breaker, Thiamine pyrophosphate riboswitches are targets for the antimicrobial compound pyrithiamine, *Chemistry & Biology* 12 (2005) 1325–1335.
- [14] S. Wang, K. A. White, Riboswitching on RNA virus replication, *Proceedings of the National Academy of Sciences* 104 (2007) 10406–10411.
- [15] P. Ketzer, J. K. Kaufmann, S. Engelhardt, S. Bossow, C. von Kalle, J. S. Hartig, G. Ungerechts, D. M. Nettelbeck, Artificial riboswitches for gene expression and replication control of DNA and RNA viruses, *Proceedings of the National Academy of Sciences* 111 (2014) E554–E562.
- [16] M. Santillán, M. C. Mackey, Dynamic behaviour of the B12 riboswitch, *Physical Biology* 2 (2005) 29.
- [17] C. L. Beisel, C. D. Smolke, Design principles for riboswitch function, *PLoS Computational Biology* 5 (2009) e1000363.
- [18] I. L. Hofacker, C. Flamm, C. Heine, M. T. Wolfinger, G. Scheuermann, P. F. Stadler, Barmap: RNA folding on dynamic energy landscapes, *RNA* 16 (2010) 1308–1316.
- [19] M. T. Wolfinger, C. Flamm, I. L. Hofacker, Efficient computation of co-transcriptional rna-ligand interaction dynamics, *Methods* 143 (2018) 70–76.
- [20] Wolfram Research, Inc., *MATHEMATICA*, v. 11.2, 2017. Champaign, IL, 2018.
- [21] J. Ryals, R. Little, H. Bremer, Temperature dependence of RNA synthesis parameters in *escherichia coli*, *Journal of Bacteriology* 151 (1982) 879–887.
- [22] E. A. Abbondanzieri, J. W. Shaevitz, S. M. Block, Picocalorimetry of transcription by RNA polymerase, *Biophysical Journal* 89 (2005) L61–L63.
- [23] A. Reining, S. Nozinovic, K. Schlepckow, F. Buhr, B. Fürtig, H. Schwalbe, Three-state mechanism couples ligand and temperature sensing in riboswitches, *Nature* 499 (2013) 355.
- [24] D. R. Cox, H. D. Miller, *The Theory of Stochastic Processes*, Chapman & Hall, London, 1965.