

RESEARCH

Open Access



# Tractable RNA–ligand interaction kinetics

Felix Kühnl<sup>1</sup>, Peter F. Stadler<sup>1,2,3,4,5,6,7</sup> and Sebastian Will<sup>1,4\*</sup>

From 12th International Symposium on Bioinformatics Research and Applications (ISBRA 2016)  
Minsk, Belarus. 5-8 June 2016

## Abstract

**Background:** The binding of small ligands to RNA elements can cause substantial changes in the RNA structure. This constitutes an important, fast-acting mechanism of ligand-controlled transcriptional and translational gene regulation implemented by a wide variety of riboswitches. The associated refolding processes often cannot be explained by thermodynamic effects alone. Instead, they are governed by the kinetics of RNA folding. While the computational analysis of RNA folding can make use of well-established models of the thermodynamics of RNA structures formation, RNA–RNA interaction, and RNA–ligand interaction, kinetic effects pose fundamentally more challenging problems due to the enormous size of the conformation space. The analysis of the combined process of ligand binding and structure formation even for small RNAs is plagued by intractably large state spaces. Moreover, the interaction is concentration-dependent and thus is intrinsically non-linear. This precludes the direct transfer of the strategies previously used for the analysis of RNA folding kinetics.

**Results:** In our novel, computationally tractable approach to RNA–ligand kinetics, we overcome the two main difficulties by applying a gradient-based coarse graining to RNA–ligand systems and solving the process in a pseudo-first order approximation. The latter is well-justified for the most common case of ligand excess in RNA–ligand systems. We present the approach rigorously and discuss the parametrization of the model based on empirical data. The method supports the kinetic study of RNA–ligand systems, in particular at different ligand concentrations. As an example, we apply our approach to analyze the concentration dependence of the ligand response of the rationally designed, artificial theophylline riboswitch RS3.

**Conclusion:** This work demonstrates the tractability of the computational analysis of RNA–ligand interaction. Naturally, the model will profit as more accurate measurements of folding and binding parameters become available. Due to this work, computational analysis is available to support tasks like the design of riboswitches; our analysis of RS3 suggests strong co-transcriptional effects for this riboswitch. The method used in this study is available online, cf. Section “Availability of data and materials”.

**Keywords:** RNA secondary structure prediction, RNA interaction kinetics, RNA–ligand interaction, Riboswitches

## Background

Riboswitches enable the specific response to the presence of ligands by transcriptional or translational control of gene expression. Their ability to switch genes on or off depending on small molecules such as theophylline or tetracycline makes them valuable biotechnological tools.

The design of tailored riboswitches for specific applications and advanced control logic is therefore an attractive endeavor in synthetic biology [1]. A riboswitch can be understood as the composition of its aptamer and its actuator domain. It senses the ligand by binding it to a binding pocket of the aptamer domain; this influences the conformations of the actuator domain and thereby leads to a measurable response to ligand binding, e. g. by terminating transcription (OFF-switch) or suppressing the terminator hairpin (ON-switch).

\*Correspondence: will@bioinf.uni-leipzig.de

<sup>1</sup>Department of Computer Science and Interdisciplinary Center for Bioinformatics, University Leipzig, Härtelstr. 16-18, D-04107, Leipzig, Germany

<sup>4</sup>Department Theoretical Chemistry, University Vienna, Währingerstr. 17, A-1090 Wien, Austria

Full list of author information is available at the end of the article

The computational design of artificial riboswitches requires a sufficiently accurate model of the ligand binding process and the structural response of the RNA to ligand binding. The equilibrium thermodynamics of RNA–ligand binding has been studied for RNA–RNA interactions, e. g. in [2, 3], and for small molecule binding in RNA–ligand [4]. As in the case of single molecule RNA folding, purely thermodynamic models are sometimes insufficient because they disregard the dynamics of the process. This can cause dramatic mis-predictions. Various approaches have analyzed the kinetics of single molecule RNA folding [5–8]. For tractability, the continuous process is decomposed into elementary steps, simplified based on heuristic assumptions, and/or approximated by a coarse-grained process.

Wolfinger et al. [7] present a coarse-graining approach to approximate RNA folding. Following e. g. [6], they analyze the folding process on the energy landscape of conformations, i. e. secondary structures,  $R_i$  of an RNA  $R$ . Conformation change is modeled by elementary moves (base pair insertion or deletion) endowed with reaction rates that follow the Arrhenius rule and thus depend on the energy barrier between the source and target conformations. In the approximation of RNA secondary structures, activation energies for opening/closing of single base pairs are approximately constant. The energy barrier thus effectively depends only on the energy difference between source and target [7]. This defines a Markov Process on the state space of all secondary structures, which is too large to make it possible to analyze it by diagonalizing the corresponding rate matrix. To effectively reduce the state space, [7] combine states into basins that consist of all conformations that are connected to the same local minimum by their gradient walk on the energy landscape. Since gradient walks connect states to their lowest energy neighbors, they correspond to the fastest folding paths from a state into a local minimum. This provides the rationale for approximating the full process by the *macroprocess* on gradient basin *macrostates*, which are assumed to be equilibrated. Consequently, the rates between the macrostates are canonically derived as weighted sums of *microrates* of the original process. By employing this heuristic approach, the size of the conformation space of smaller RNA molecules of up to a hundred nucleotides is typically reduced to just a few thousand macrostates. For example, in our analysis of the 81 nucleotide long riboswitch RS3 [1], 11.4 millions of secondary structures are mapped to 1133 macrostates. The macroprocess is finally solved by diagonalization. In our approach we re-use ideas of this coarse-graining, which also allows us to re-use several tools for single-molecule RNA kinetics (*RNAsubopt* [9], *barriers* [10], *treekin* [7]).

### RNA–ligand interaction model

We are going to describe a reaction system of the RNA  $R$  (given by its sequence of nucleotides  $\{A, C, G, U\}$ ) with the ligand  $L$  at the level of RNA and binding complex conformations, such that we can study the kinetics of association, dissociation, and conformation changes. For simplicity, we assume that there is only a single ligand conformation (also denoted  $L$ ). In the same way as a single RNA molecule transitions between various conformations during the folding process, the complex of RNA and ligand adopts different conformations  $LR_i$ . Importantly, only a subset of the RNA conformations binds the ligand. The part of the total state space that corresponds to the RNA–ligand complexes is therefore isomorphic to a subset of the state space of the free RNA molecule. Thus, our system of consideration consists of the reactions

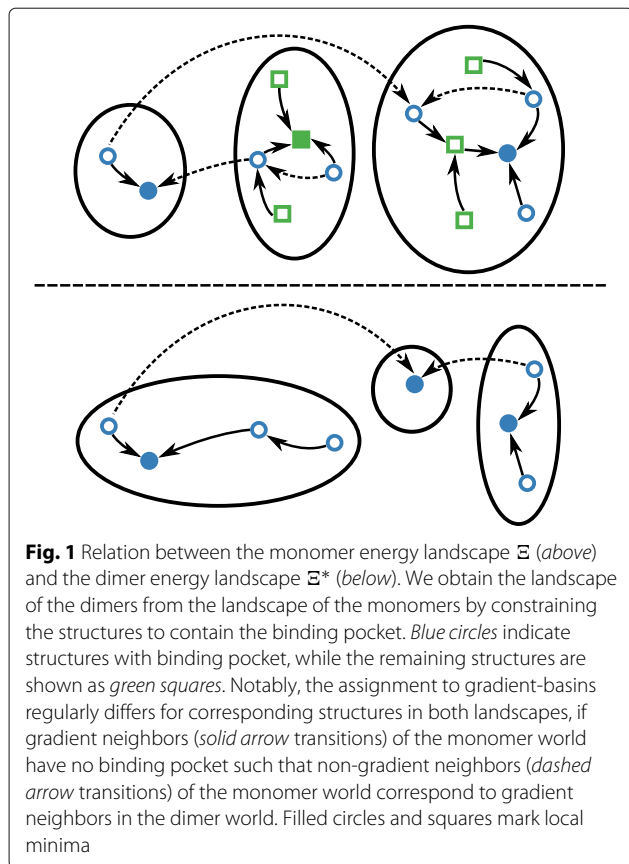


for all  $i, j \in \{1, \dots, N\}$ ,  $i \neq j$ . According to the rate laws for elementary reactions, the rates of each of these reactions depend on specific rate constants and the concentrations of the molecules. The reactions 1 and 4 only have non-zero rate constants, if the RNA conformations  $R_i$  and  $R_j$  are related by an elementary move such as the insertion or deletion of a base pair. Moreover,  $L$  and  $R_i$  can interact only for the subset of the  $R_i$  that form an appropriate binding pocket; otherwise, the complex  $LR_i$  is deemed unstable and thus excluded from the model. Since RNA conformations correspond to RNA secondary structures, the energies of monomer states can be calculated from the Turner energy model [11]. For dimer states, we add the aptamer–ligand-specific binding energy. For the exemplary studied riboswitch RS3, this energy can be derived from the empirical dissociation constant [1]. Finally, we derive the rate constants as Metropolis rates with appropriate pre-exponential factors that can be estimated from empirical rates. Note that the rates of base pair opening  $k^-$  and closing  $k^+$  are directly related by the energy change  $\Delta G$  due to the closing. Concretely,  $k^-/k^+ = \exp(\Delta G/RT)$  for  $\Delta G < 0$ . Experimental values are available for the zippering rate, which corresponds to the rate of closing the last hairpin in a helix. A careful analysis in [12] yields a value in the range  $4.7 \cdot 10^7$  to  $10^9 \text{s}^{-1}$  roughly consistent with earlier estimates [13–15]. In principle, a kinetic constant can be derived for the closing of first base pair in a loop from a worm-like chain model [12, 16]; following earlier work on RNA kinetic models [7], we use here a single kinetic parameter  $k^+$  for all base pairs. An empirical rate of one specific theophylline aptamer association was

reported as  $600 \text{ M}^{-1}\text{s}^{-1}$  [17], which may serve as rough estimate for comparable systems. Note that [17] measured the macroscopic *apparent rate* that depends on the rate of dimerization as well as the rate of refolding into structures with theophylline binding pocket.

While the Reactions 1, 3, and 4 are of first order, the second order association in Reaction 2 introduces non-linearity into the system. Assuming ligand excess, which is a very plausible assumption for small molecular ligands, however, it is possible to devise a *pseudo-first order approximation* of the system.

Even if the reaction equations above appropriately model the RNA–ligand interaction, this system is still computationally intractable for typical riboswitch sizes. As a remedy, we construct a coarse-grained process based on the separate gradient-basins for the monomer and dimer-states. The monomer states with suitable binding pocket are connected to dimer states, cf. Fig. 1. Importantly, there is no direct mapping from monomer macrostates to dimer macrostates of our coarse-grained system because conformations without binding pockets are absent from the “dimer world”. The upper basin in the “monomer world” of Fig. 1 is subdivided into two basins in the dimer world; conversely, the middle and lower



monomer basins correspond to a single basin of the dimer world.

### Contributions

We start by elaborating the general macroprocess of RNA–ligand interaction, based on gradient basin macrostates, and derive the corresponding rate constants. This contributes an original description of coarse-grained interaction processes; it is the first fundamental prerequisite for our tractable RNA–ligand kinetics approach. Furthermore, we leverage that a wide spectrum of biological RNA–ligand systems operate under strong ligand excess, justifying the pseudo-first order approximation. On these grounds, we establish the first analytical approach for RNA–ligand interaction kinetics. Based on solving the master equation of the interaction process, this enables the computation of time-dependent macrostate probabilities. Finally, we study the kinetics of the artificially designed riboswitch RS3 [1] interacting with theophylline. We analyze the system at different concentrations and present results that strongly suggest co-transcriptional effects.

### Methods

We consider the fixed interaction system of the RNA  $R$  and the ligand  $L$ . Let  $X$  denote the set of all *monomer microstates*,  $X = \{R_i \mid i = 1, \dots, N\}$ ; in our setting, the  $R_i$  are the secondary structures of a given RNA sequence. The subset  $X^+ \subseteq X$  comprises the conformations that can bind the ligand. Here  $X^+$  contains all states with a specific binding pocket. Furthermore, define  $X^*$  as the set of *dimer microstates*  $LR_i$ ,  $X^* = \{LR_i \mid R_i \in X^+\} \subseteq \{LR_i \mid i = 1, \dots, N\}$ .

A dimer microstate  $LR_i \in X^*$  has the energy  $E(R_i) + \theta$ , where  $\theta < 0$  denotes the *binding energy* of  $R$  and  $L$ . The *inverse temperature* is  $\mathfrak{b} = \frac{1}{RT}$ , where  $T$  is the *absolute temperature* and  $R$  is the *universal gas constant*. For a set  $S \subseteq X \cup X^*$  of microstates, let  $Z[S] := \sum_{x \in S} \exp(-\mathfrak{b}E(x))$  denote the partition function of  $S$ . The *probability of a microstate*  $x$  in  $S$  is then given by

$$\Pr[x \mid S] := \begin{cases} \exp(-\mathfrak{b}E(x)) / Z[S] & \text{if } x \in S, \\ 0 & \text{otherwise.} \end{cases}$$

Let  $x, y \in X \cup X^*$  be microstates. The *microrate constant* from  $x$  to  $y$  is denoted  $k(x \rightarrow y)$  (or  $k(y \leftarrow x)$ ). On microstates  $x, y$ , we define the symmetric *neighborhood relation*  $\mathcal{N}$  such that  $x \mathcal{N} y$  holds if and only if  $x$  and  $y$  have a distance of exactly one elementary move.

For  $x \mathcal{N} y$ , let

$$E_{xy}^\Delta := \max\{E(x), E(y)\} - E(x)$$

be the activation energy of the transition  $x \rightarrow y$  as defined by the Metropolis rule. Accordingly, the microrate constants for distinct states  $x, y \in X \cup X^*$ , where  $x \mathcal{N} y$ , are defined as

$$k(x \rightarrow y) := A(x \rightarrow y) \exp(-bE_{xy}^\Delta),$$

otherwise, define  $k(x \rightarrow y) := 0$ .  $A(x \rightarrow y)$  denotes the reaction-specific *pre-exponential factor*. For our purposes, we assume that this factor depends only on the type of reaction and the factors for conformation change in monomers and dimers are equal. Thus, we distinguish the factors  $A_a$  for association,  $A_d$  for dissociation, and  $A_R$  for conformation changes of the RNA secondary structure. As we will show later,  $A_a = A_d$  due to detailed balance.

Denote the *powerset* of a set  $S$  by  $\mathcal{P}(S)$ . A *monomer (or dimer) macrostate* is a set of monomer (or dimer) microstates, i. e. an element of  $\mathcal{P}(X)$  ( $\mathcal{P}(X^*)$ ). We denote the (*macro*)rate constant from macrostate  $\alpha$  to  $\beta$  by  $r(\alpha \rightarrow \beta)$  (or  $r(\beta \leftarrow \alpha)$ ). Macrorate constants are defined by summing over microrate constants and their respective state probabilities, i. e.

$$r(\alpha \rightarrow \beta) := \sum_{x \in \alpha, y \in \beta} \Pr[x | \alpha] \cdot k(x \rightarrow y).$$

We emphasize that we use the term macrostates freely to denote general sets of microstates. Only when we introduce specific partitions of the microstates into macrostates, it makes sense to distinguish *represented* macrostates of our specific coarse-grained system from other sets of microstates.

## Results and discussion

### Macrostate kinetics of RNA–ligand interaction

For a microstate  $x \in X^+$ , we denote its corresponding dimer microstate (after binding to  $L$ ) by  $Lx$ , i. e. for  $x = R_i$ ,  $Lx = LR_i$ . This notation is raised to sets of microstates by defining  $L\alpha := \{Lx | x \in \alpha\}$ . Lemma 1 below asserts that the rate constants between dimer microstates and macrostates can be computed exactly like rate constants of monomer states.

**Lemma 1** For  $x, y \in X^+$ ,  $k(Lx \rightarrow Ly) = k(x \rightarrow y)$ . Furthermore,  $\Pr[Lx | L\alpha] = \Pr[x | \alpha]$  holds for all  $\alpha \in \mathcal{P}(X^+)$ . Finally,  $r(L\alpha \rightarrow L\beta) = r(\alpha \rightarrow \beta)$ , for all macrostates  $\alpha, \beta \in \mathcal{P}(X^+)$ .

*Proof* The individual claims follow easily from the definitions. If  $(x, y) \notin \mathcal{N}$ ,  $k(x \rightarrow y) = 0 = k(Lx \rightarrow Ly)$ , so assume  $x \mathcal{N} y$ . Since

$$\begin{aligned} E_{LxLy}^\Delta &= \max\{E(Lx), E(Ly)\} - E(Lx) \\ &= \max\{E(x) + \theta_L, E(y) + \theta_L\} \\ &\quad - (E(x) + \theta_L) \\ &= \max\{E(x), E(y)\} + \theta_L \\ &\quad - E(x) - \theta_L \\ &= E_{xy}^\Delta \end{aligned}$$

holds,

$$\begin{aligned} k(Lx \rightarrow Ly) &= A(Lx \rightarrow Ly) \exp(-bE_{LxLy}^\Delta) \\ &= A(x \rightarrow y) \exp(-bE_{xy}^\Delta) \\ &= k(x \rightarrow y). \end{aligned}$$

Furthermore,

$$\begin{aligned} \Pr[Lx | L\alpha] &= \frac{\exp(-bE(Lx))}{Z[L\alpha]} \\ &= \frac{\exp(-b[E(x) + \theta_L])}{\sum_{x \in \alpha} \exp(-b[E(x) + \theta_L])} \\ &= \frac{\exp(-b\theta_L) \exp(-bE(x))}{\exp(-b\theta_L) \sum_{x \in \alpha} \exp(-bE(x))} \\ &= \Pr[x | \alpha]. \end{aligned}$$

Finally,

$$\begin{aligned} r(L\alpha \rightarrow L\beta) &= \sum_{\substack{Lx \in L\alpha \\ Ly \in L\beta}} \Pr[Lx | L\alpha] \cdot k(Lx \rightarrow Ly) \\ &= \sum_{\substack{x \in \alpha \\ y \in \beta}} \Pr[x | \alpha] \cdot k(x \rightarrow y) \\ &= r(\alpha \rightarrow \beta). \end{aligned}$$

□

The microrate constant from monomer to dimer states is constant, whereas the back rate depends on the binding energy  $\theta$ .

**Lemma 2** (Association and dissociation microrate constants) For  $x \in X^+$ , the rate of association is  $k(x \rightarrow Lx) = A_a$ , while the dissociation rate is  $k(Lx \rightarrow x) = A_d \exp(b\theta)$ . All other rates between monomer and dimer microstates are 0.

*Proof* By Metropolis rule, for  $x \in X^+$ ,

$$\begin{aligned} k(x \rightarrow Lx) &= A_a \exp(-bE_{xLx}^\Delta) \\ &= A_a \exp(-b[E(x) - E(x)]) \\ &= A_a, \end{aligned}$$

since  $E_{xLx}^\Delta = \max\{E(x), E(Lx)\} - E(x)$  and, additionally,  $E(Lx) = E(x) + \theta_L \leq E(x)$ . Analogously, for the inverse microrate,

$$\begin{aligned} k(Lx \rightarrow x) &= A_d \exp(-bE_{xLx}^\Delta) \\ &= A_d \exp(-b[E(x) - E(Lx)]) \\ &= A_d \exp(b\theta_L). \end{aligned}$$

□

The association (dissociation) microrates due to Lemma 2 induce corresponding macrorates, which additionally depend on the probability of the associable (dissociable) microstates in the source macrostate, respectively (Lemma 3).

**Lemma 3** (Association and dissociation macrorate constants) *For arbitrary  $\alpha \in \mathcal{P}(X)$  and  $\beta \in \mathcal{P}(X^+)$ , equation  $r(\alpha \rightarrow L\beta) = A_a \frac{Z[\alpha \cap \beta]}{Z[\alpha]}$  holds. Additionally,  $r(L\beta \rightarrow \alpha) = A_d \frac{Z[\alpha \cap \beta]}{Z[\beta]} \cdot \exp(b\theta)$ .*

*Proof* Let  $\alpha \in \mathcal{P}(X)$  and  $\beta \in \mathcal{P}(X^+)$ . Thus

$$\begin{aligned} r(\alpha \rightarrow L\beta) &= \sum_{\substack{x \in \alpha \\ Ly \in L\beta}} \Pr[x | \alpha] \cdot k(x \rightarrow Ly) \\ &= \sum_{x \in \alpha \cap \beta} \Pr[x | \alpha] \cdot k(x \rightarrow Lx) \\ &= A_a \frac{Z[\alpha \cap \beta]}{Z[\alpha]} \end{aligned}$$

and

$$\begin{aligned} r(L\beta \rightarrow \alpha) &= \sum_{\substack{Lx \in L\beta \\ y \in \alpha}} \Pr[Lx | L\beta] \cdot k(Lx \rightarrow y) \\ &= \sum_{x \in \alpha \cap \beta} \Pr[Lx | L\beta] \cdot k(Lx \rightarrow x) \\ &= A_d \frac{Z[\alpha \cap \beta]}{Z[\beta]} \cdot \exp(b\theta_L). \end{aligned}$$

□

### A tractable model under ligand excess

For our coarse-grained RNA–ligand interaction process, we partition the monomer microstates  $X$  and the dimer microstates  $X^*$  into sets of macrostates  $\Xi$  and  $\Xi^*$ , respectively. For the theoretical discussion, we require only that  $\Xi$  and  $\Xi^*$  are partitions of the respective sets  $X$  and  $X^*$ . Later, in our application, we are going to define macrostates as gradient basins (within their respective component).

We denote the monomer macrostates (in  $\Xi$ ) by  $\alpha_1, \dots, \alpha_n$  and the dimer macrostates (in  $\Xi^*$ ) by  $\beta_1, \dots, \beta_m$ . Since—by model assumption—the ligand is in large excess, the change of the ligand concentration  $[L]$  is essentially negligible in relation to the change of RNA concentrations. Formally, we assume  $d/dt[L] = 0$ , i. e. at all times  $[L] = l_0$ , for the initial ligand concentration  $l_0$ . The

change of RNA monomer and RNA–ligand dimer concentrations over time is described by a system of ordinary differential equations (ODEs) corresponding to Reactions (1)–(4).

Following the first-order rate laws, Reaction (1) causes  $n^2 - n$  flows  $r(\alpha_i \rightarrow \alpha_j)[\alpha_i]$  from  $\alpha_i$  to  $\alpha_j$  ( $1 \leq i, j \leq n, i \neq j$ ); Reaction (4)  $m^2 - m$  flows  $r(\beta_i \rightarrow \beta_j)[\beta_i]$  from  $\beta_i$  to  $\beta_j$  ( $1 \leq i, j \leq m, i \neq j$ ); and Reaction (3)  $n \cdot m$  flows  $r(\beta_i \rightarrow \alpha_j)[\beta_i]$  from  $\beta_i$  to  $\alpha_j$  ( $1 \leq i \leq m$  and  $1 \leq j \leq n$ ). In contrast to these simple first-order transitions, the state changes due to Reaction 2 follow second-order rate laws contributing the  $n \cdot m$  flows  $r(\alpha_i \rightarrow \beta_j)[L][\alpha_i]$  from  $\alpha_i$  to  $\beta_j$  ( $1 \leq i \leq n$  and  $1 \leq j \leq m$ ). Without the assumption  $d/dt[L] = 0$ , the rate would depend on two variable concentrations, causing the system to be non-linear. However, by our assumption, the concentration  $[L]$  is constant.

**The system of ODEs.** The change of concentrations is now described by summing over single contributions:

$$\begin{aligned} \frac{d}{dt}[\alpha_i] &= \sum_{\substack{1 \leq k \leq n \\ k \neq i}} r(\alpha_k \rightarrow \alpha_i)[\alpha_k] \\ &\quad + \sum_{1 \leq k \leq m} r(\beta_k \rightarrow \alpha_i)[\beta_k] \\ &\quad - \sum_{\substack{1 \leq k \leq n \\ k \neq i}} r(\alpha_i \rightarrow \alpha_k)[\alpha_i] \\ &\quad - \sum_{1 \leq k \leq m} r(\alpha_i \rightarrow \beta_k)[L][\alpha_i] \end{aligned}$$

for  $i = 1, \dots, n$ , and

$$\begin{aligned} \frac{d}{dt}[\beta_j] &= \sum_{1 \leq k \leq n} r(\alpha_k \rightarrow \beta_j)[L][\alpha_k] \\ &\quad + \sum_{\substack{1 \leq k \leq m \\ k \neq j}} r(\beta_k \rightarrow \beta_j)[\beta_k] \\ &\quad - \sum_{1 \leq k \leq n} r(\beta_j \rightarrow \alpha_k)[\beta_j] \\ &\quad - \sum_{\substack{1 \leq k \leq m \\ k \neq j}} r(\beta_j \rightarrow \beta_k)[\beta_j] \end{aligned}$$

for  $j = 1, \dots, m$ .

We set  $\gamma := (\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_m)^T$  and define the  $(n + m) \times (n + m)$ -matrix  $R(l_0)$ . Then the entire coarse-grained system under ligand excess can be expressed by the linear ODE  $\frac{d}{dt}[\gamma] = R(l_0)[\gamma]$ , where

$$R(l_0) = \begin{pmatrix} A & C \\ l_0 \cdot D & B \end{pmatrix}$$

is constructed from four submatrices:

A is an  $n \times n$ -matrix with entries  $a_{ij} = r(\alpha_i \leftarrow \alpha_j)$  for  $1 \leq i, j \leq n, i \neq j$ . For  $1 \leq i \leq n$ ,

$$a_{ii} := - \sum_{\substack{1 \leq k \leq n \\ k \neq i}} r(\alpha_k \leftarrow \alpha_i) - \sum_{1 \leq k \leq m} l_0 r(\beta_k \leftarrow \alpha_i).$$

B is an  $m \times m$ -matrix with entries  $b_{ij} = r(\beta_i \leftarrow \beta_j)$  for  $1 \leq i, j \leq m, i \neq j$ . For  $1 \leq j \leq m$ ,

$$b_{jj} := - \sum_{1 \leq k \leq n} r(\alpha_k \leftarrow \beta_j) - \sum_{\substack{1 \leq k \leq m \\ k \neq j}} r(\beta_k \leftarrow \beta_j).$$

C is an  $n \times m$ -matrix with entries  $c_{ij} = r(\alpha_i \leftarrow \beta_j)$  for  $1 \leq i \leq n, 1 \leq j \leq m$ , and

D is an  $m \times n$ -matrix with entries  $d_{ij} = r(\beta_i \leftarrow \alpha_j)$  for  $1 \leq i \leq m, 1 \leq j \leq n$ .

### Computing RNA–ligand kinetics

The described ODE system can be solved analytically building on existing software. The entire computation pipeline consists of five major steps:

1. enumeration of the RNA’s structure space
2. computation of the gradient basins and corresponding rates for
  - (a) the monomer landscape
  - (b) the dimer landscape
3. computation of the rates between the monomer and dimer basins
4. construction of the full rate matrix  $R(l_0)$
5. integration of the linear ODE system

Since an exhaustive enumeration of the structure space is infeasible even for short RNAs, *Step 1* generates only a selected part of all possible secondary structures of the input RNA. For this work, we consider only structures up to a certain energy above the minimum free energy of the sequence as computed by `RNAsubopt` [9]. To further reduce the number of structures, only structures without any isolated base pairs are generated.

Often, the restriction to low energy structures excludes important microstates of relatively high energy such as the open RNA chain from the model. Simply adding such a structure to the system is insufficient without also including transitional structures that connect it to the remaining states. The solution for this work was to develop a heuristic algorithm to partially explore an energy landscape around a given structure of interest, flooding neighbored basins if their local minimum has a lower energy than the current one, until one reaches structures within the already explored energy band. This approach seems to be more adequate in the context of a gradient basin coarse graining than a direct path heuristic (e. g. `findPath` [6]).

In *Step 2a*, we compute the gradient basins and rates for the monomer landscape from the list of input structures using `barriers` [7] (with `minh` heuristic). For *Step 2b*, a list of all input structures that contain the binding pocket is generated with `RNAsubopt`’s constraint folding mode. This enables one to enumerate dimer structures up to a higher energy than possible for the entire landscape, ensuring the dimer world is connected. As shown in Lemma 1, the transition rates in the constrained dimer landscape are independent of the ligand’s binding energy and thus can be computed exactly like those of the monomer landscape.

In *Step 3*, the transition rates between monomer and dimer macrostates are computed based on Lemma 3 using the mapping of the monomer and dimer structures to their respective basins. For this purpose we modified `barriers` to output this information.

*Step 4* yields the full rate matrix  $R(l_0)$  for one set of pre-exponential factors and a certain ligand concentration  $l_0$  by combining the previously computed rate constants. We emphasize that we can easily compute  $R(l_0)$  for different values of  $l_0$ ,  $A_A$  and  $A_R$  without repeating the previous, more time-consuming computation steps.

Finally, in *Step 5* the system of ODEs is solved directly using the closed form  $\vec{c}(t) = \exp(tR(l_0)) \cdot \vec{c}(0)$ , where  $\vec{c}(t)$  is the vector of macrostate concentrations at time  $t$ . The exponential  $\exp(tR(l_0))$  is obtained by diagonalizing  $R(l_0)$  numerically using the tool `treekin` [7], which performs this computation efficiently.

### Parameters from empirical measurements

The binding energy  $\theta$  can be derived from an empirically measured dissociation constant  $K_d^A$  of the aptamer; e. g. in the case of theophylline, [18] measure a  $K_d^A$  of  $0.32 \mu M$  for the theophylline aptamer of *RS3*. From the macroscopic measurement, we derive the binding energy as

$$\theta = RT_A \ln \left( K_d^A \cdot \Pr[ \text{“pocket”} \mid A, T_A ] \right),$$

where  $T_A = 298K$  is the temperature of the measurement,  $R$  is the gas constant, and  $\Pr[ \text{“pocket”} \mid A, T_A ]$  denotes the equilibrium probability of the binding pocket in the aptamer at temperature  $T_A$  as calculated in the Turner energy model (cf. [1], which neglect the probability). This relation allows calculating the effective dissociation constant at temperature  $T_R$  of a theophylline riboswitch like *RS3* that contains the aptamer, due to the inverse relation

$$\begin{aligned} K_d^{RS} &= \frac{\exp\left(\frac{\theta}{RT_R}\right)}{\Pr[ \text{“pocket”} \mid RS, T_R ]} \\ &= \frac{(K_d^A \cdot \Pr[ \text{“pocket”} \mid A, T_A ])^{\frac{T_A}{T_R}}}{\Pr[ \text{“pocket”} \mid RS, T_R ]}. \end{aligned}$$

For RS3 at  $T_R = 313.15$  K,

$$\text{Pr}[\text{“pocket”} | A, T_A] \approx 0.292$$

and

$$\text{Pr}[\text{“pocket”} | RS, T_R] \approx 2.59 \cdot 10^{-11}$$

due to the pocket-constrained and unconstrained ensemble free energies in the Turner model. Thus,

$$\theta \approx RT_A \ln(0.292K_d^A) \approx -9.59 \frac{\text{kcal}}{\text{mol}}$$

and

$$K_d^{\text{RS3}} \approx \frac{(0.292K_d^A)^{\frac{T_A}{T_R}}}{2.59 \cdot 10^{-11}} \approx 7891 \text{ M.}$$

For relating the rates of the different reaction types, one needs to estimate the pre-exponential factors of all reactions. Commonly, one assumes constant factors for each type of reaction. In reasonable approximation, we furthermore equate the factors for monomer and dimer conformation changes.

Given the apparent association rate  $A_a^m$  (which we assume to equal the macroscopic pre-exponential factor of dimerization), one can bound the microscopic pre-exponential factor  $A_a$ . If we assume that refolding is much slower than dimerization, then  $A_a^m$  is a product of the microrate and the equilibrium probability of the binding pocket. Conversely, if we assume the refolding to be much faster, than  $A_a^m$  directly measures the dimerization microrate. Thus,

$$A_a^m \leq A_a \leq A_a^m \cdot \text{Pr}[\text{“pocket”} | \text{aptamer}]^{-1}.$$

In the case of theophylline,

$$\text{Pr}[\text{“binding pocket”} | \text{aptamer}] \approx 1$$

and consequently,  $A_a \approx A_a^m$ .

Finally, the pre-exponential factor for dissociation  $A_d$  equals  $A_a$ . This is a consequence of detailed balance of the dimerization reaction, i. e.

$$k(R_i \rightarrow LR_i) \text{Pr}[R_i] = k(LR_i \rightarrow R_i) \text{Pr}[LR_i],$$

which implies

$$\begin{aligned} A_a \text{Pr}[R_i] &= A_d \exp(b\theta) \text{Pr}[R_i] \exp(-b\theta) \\ &= A_d \text{Pr}[R_i]. \end{aligned}$$

### Empirical results

We apply our system to demonstrate the effect of changes in ligand concentrations to the interaction of the designed ON-switch RS3 from [1] with the ligand theophylline. Using our prototypical software, we precompute the macroprocess for RS3 including rate constants in several hours. As noted above, this yields 1133 macrostates in total, 22 of which are dimer states representing RNA

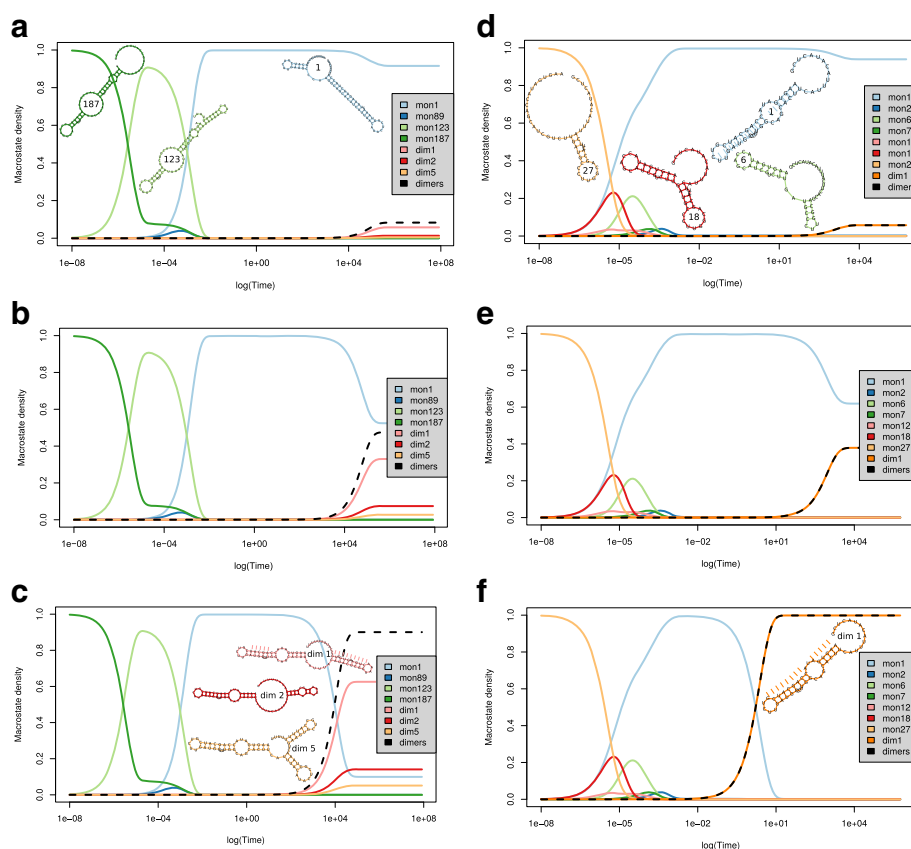
molecules bound to the ligand. From the technical point of view, our system is described by 1133 coupled differential equations for the states  $\alpha_1, \dots, \alpha_{1111}$  through  $\beta_1, \dots, \beta_{22}$ . Subsequently, we compute kinetics for each combination of concentrations and pre-exponential factors within seconds (on a Core i5-750 @  $4 \times 2.67$ GHz). Figure 2 summarizes our results; each subfigure plots the probabilities of prominent monomer and dimer states over time. In addition, the minimum energy secondary structures is shown for the most important macrostates. It serves as a suitable representative of the macrostate's ensemble of structures. In this sense it provides a useful, coarse-grained picture of the most likely refolding paths. We set the pre-exponential factors to the estimations  $A_R = 10^6 \text{s}^{-1}$  and  $A_a = 600 \text{M}^{-1} \text{s}^{-1}$  as described before. This allows interpreting the time and ligand concentrations in concrete units and relates the speed of folding and dimerization.

Figure 2a–c show the results for ligand concentrations  $10^4$  M,  $10^5$  M, and  $10^6$  M. In the RS3 riboswitch, the aptamer domain is fused to a rho-independent terminator at the 3'-end. Thus, during transcription the aptamer is available shortly before the strong terminator stem can be formed and then dominates the entire structure ensemble. Therefore, we study a partially transcribed riboswitch RS3 that is shortened by the 3'-half of the terminator stem and the 3' poly-U stretch. The kinetics of the shortened riboswitch are shown for concentrations of  $10^{-7}$  M,  $10^{-6}$  M, and  $10^{-3}$  M in respective Fig. 2d–f. Note that the time scales for interaction of RS3 with theophylline are in accordance with the computed dissociation constant  $K_d^{\text{RS3}}$ , which implies that the monomer and dimer concentrations are balanced at about  $10^4$  M ligand concentration. This extreme concentration suggests that the riboswitch would be non-functional without further, probably co-transcriptional, effects. This is a plausible hypothesis since RS3 was designed to regulate at the transcriptional level.

The estimated rates are derived from a small number of empirical measurements at different conditions, such as ion concentrations (100 mM NaCl in [12], 5 mM  $\text{MgCl}_2$  and 0.5 M NaCl in [18], no  $\text{Mg}^{2+}$  and 100 mM NaCl in [17]), temperatures, and actual sequences; hence they are not directly comparable. Nevertheless, they provide reasonable ball park estimates, because we observed that the qualitative behavior of the system is robust against variations of these parameters by several orders of magnitude.

### Conclusions

Several refinements of the model remain for future research. Most importantly, the assumption of only a single binding motif is rather stringent. In general, one would like to support multiple binding sites with different



**Fig. 2** Kinetics plots showing the probabilities of prominent monomer and dimer states (*y-axis*) over time in seconds (*x-axis*) at a RNA folding rate of  $10^6 s^{-1}$  and a dimerization rate of  $600 M^{-1} s^{-1}$ . Additionally, we visualize the the most prominent macrostates by their local minimum structures, which enables tracking of their coarse-grained refolding. **a–c** Complete riboswitch RS3 at concentrations  $10^4$  M (**a**),  $10^5$  M (**b**), and  $10^5$  M (**c**). **d–f** Partially transcribed riboswitch RS3 (without 3'-half of terminator stem) at concentrations  $10^{-7}$  M (**d**),  $10^{-6}$  M (**e**), and  $10^{-3}$  M (**f**). Note that since subfigures **a–c** are based on exactly the same landscapes, they share the same macrostates (e. g., mon1 in **a** and mon1 in **b** are equal). As well, this holds among subfigures **d–f**. However, across the two groups of subfigures, macrostates are not comparable (e. g., mon1 of **a**  $\neq$  mon1 of **d**), since the landscapes differ

binding energies. Our model can be naturally generalized to such scenarios by introducing multiple “dimer worlds” corresponding to different binding motifs. Furthermore, some ligands, such as  $Mg^{2+}$  have multiple binding sites. The current implementation of the Arrhenius approximation of the RNA folding kinetics, finally, is quite simplistic, using only a single kinetic prefactor for all structural rearrangements. A refined model would presumably distinguishing constants for nucleation, stack extension, base pair sliding, and loop pinching. Moreover, in particular transcriptional riboswitches, which operate temporally coupled with the progressive transcription of the RNA, will be influenced by this kinetic interplay. It is well known that RNA chains commonly change their optimal structure while growing during transcription. Consequently, the RNAs refold during the process of transcription [8]. The framework presented here can be extended to co-transcriptional interaction analysis. However, this

will require additional experimental measurements to calibrate the parameters of the model to properly relate the different “reaction” speeds. In particular, this entails accurate measurements of the thermodynamic parameters for the ligand binding and of the kinetic prefactors of folding and dimerization as well as the speed of transcription.

#### Abbreviations

ODE; Ordinary differential equation

#### Acknowledgements

A two-page abstract of this work appeared at ISBRA'16 [19].

#### Funding

This work is supported by the German Federal Ministry of Education and Research (BMBF; support code 031A538B) within the German Network for Bioinformatics Infrastructure “de.NBI” and by the German Research Foundation (DFG; grant STA 850/15-1). The authors acknowledge funding support for publication charges from the German Research Foundation (DFG) and Universität Leipzig within the program of Open Access Publishing. The authors declare that the funding body has not been influencing the design or conclusion of this study.



**Availability of data and materials**

The method used in this study is available as free software at [www.bioinf.uni-leipzig.de/~felix/software/RLIkin](http://www.bioinf.uni-leipzig.de/~felix/software/RLIkin).

**About this supplement**

This article has been published as part of *BMC Bioinformatics* Volume 18 Supplement 12, 2017: Selected articles from the 12th International Symposium on Bioinformatics Research and Applications (ISBRA-16): [bioinformatics](http://bioinformatics.biomedcentral.com/articles/supplements/volume-18-supplement-12). The full contents of the supplement are available online at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-18-supplement-12>.

**Authors' contributions**

FK implemented the method and conducted experiments. FK, PFS, and SW designed the algorithms and experiments. All authors wrote the manuscript. All authors read and approved the final manuscript.

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

**Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Author details**

<sup>1</sup>Department of Computer Science and Interdisciplinary Center for Bioinformatics, University Leipzig, Härtelstr. 16-18, D-04107, Leipzig, Germany. <sup>2</sup>MPI for Mathematics in the Sciences, Inselstr. 22, D-04103 Leipzig, Germany. <sup>3</sup>FHI Cell Therapy and Immunology, Perlickstr. 1, D-04103 Leipzig, Germany. <sup>4</sup>Department Theoretical Chemistry, University Vienna, Währingerstr. 17, A-1090 Wien, Austria. <sup>5</sup>Bioinformatics and Computational Biology Research Group, Währingerstr. 17, A-1090 Wien, Austria. <sup>6</sup>RTH, University Copenhagen, Grønnegårdsvej 3, 1870 Frederiksberg C, Denmark. <sup>7</sup>Santa Fe Institute, 1399 Hyde Park Rd., Santa Fe, NM, USA.

Published: 16 October 2017

**References**

1. Wachsmuth M, Findeiss S, Weissheimer N, Stadler PF, Mörl M. De novo design of a synthetic riboswitch that regulates transcription termination. *Nucleic Acids Res.* 2013;41(4):2541–51.
2. Dimitrov RA, Zuker M. Prediction of hybridization and melting for double-stranded nucleic acids. *Biophys J.* 2004;87(1):215–26.
3. Bernhart SH, Tafer H, Mückstein U, Flamm C, Stadler PF, Hofacker IL. Partition function and base pairing probabilities of RNA heterodimers. *Algorithms Mol Biol.* 2006;1(1):3.
4. Borujeni AE, Mishler DM, Wang J, Huso W, Salis HM. Automated physics-based design of synthetic riboswitches from diverse RNA aptamers. *Nucleic Acids Res.* 2015;44(1):1–13. Available from: <http://dx.doi.org/10.1093/nar/gkv1289>.
5. Mann M, Kucharik M, Flamm C, Wolfinger MT. Memory efficient RNA energy landscape exploration. *Bioinformatics.* 2014;30(18):2584–91.
6. Flamm C, Fontana W, Hofacker IL, Schuster P. RNA folding at elementary step resolution. *RNA.* 2000;6(3):325–38.
7. Wolfinger MT, Svrcek-Seiler WA, Flamm C, Hofacker IL, Stadler PF. Efficient computation of RNA folding dynamics. *J Phys A: Math General.* 2004;37(17):4731–41. Available from <http://stacks.iop.org/0305-4470/37/4731>.
8. Hofacker IL, Flamm C, Heine C, Wolfinger MT, Scheuermann G, Stadler PF. BarMap: RNA folding on dynamic energy landscapes. *RNA.* 2010;16(7):1308–16.
9. Lorenz R, Bernhart SH, Höner Zu Siederdisen C, Tafer H, Flamm C, Stadler PF, et al. ViennaRNA Package 2.0. *Algorithms Mol Biol.* 2011;6:26.
10. Flamm C, Hofacker IL, Stadler PF, Wolfinger MT. Barrier trees of degenerate landscapes. *Zeitschrift für Physikalische Chemie.*

2002;216(2/2002):. Available from: <http://dx.doi.org/10.1524/zpch.2002.216.2.155>.

11. Turner DH, Mathews DH. NNDB: the nearest neighbor parameter database for predicting stability of nucleic acid secondary structure. *Nucleic Acids Res. 2009;38(Database):D280–2*. Available from: <http://dx.doi.org/10.1093/nar/gkp892>.
12. Kuznetsov SV, Ansari A. A Kinetic Zipper Model with Intrachain interactions applied to nucleic acid hairpin folding kinetics. *Biophys J.* 2012;102(1):101–11. Available from: <http://dx.doi.org/10.1016/j.bpj.2011.11.4017>.
13. Pörschke D. Model Calculations on the Kinetics of Oligonucleotide Double Helix Coil Transitions. Evidence for a fast chain sliding reaction. *Biophys Chem.* 1974;2:83–96.
14. Cocco S, Marko JF, Monasson R. Slow nucleic acid unzipping kinetics from sequence-defined barriers. *Eur Phys J E Soft Matter.* 2003;10:153–61.
15. Zhang W, Chen SJ. RNA hairpin-folding kinetics. *Proc Natl Acad Sci USA.* 2002;99:1931–6.
16. Toan NM, Morrison G, Hyeon C, Thirumalai D. Kinetics of loop formation in polymer chains. *J Phys Chem B.* 2008;112:6094–106.
17. Latham MP, Zimmermann GR, Pardi A. NMR (Chemical Exchange as a Probe for Ligand-Binding Kinetics) in a Theophylline-Binding RNA Aptamer. *J Am Chem Soc.* 2009;131(14):5052–53. Available from: <http://dx.doi.org/10.1021/ja900695m>.
18. Jenison RD, Gill SC, Pardi A, Polisky B. High-resolution molecular discrimination by RNA. *Science.* 1994;263(5152):1425–9.
19. Kühnl F, Stadler PF, Will S. Tractable Kinetics of RNA-Ligand Interaction In: Bourgeois A, Skums P, Wan X, Zelikovsky A, editors. *Bioinformatics Research and Applications: 12th International Symposium*. vol. 9683 of *Lecture Notes in Bioinformatics*. Berlin: Springer International Publisher; 2016. p. 337–8.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

