The Folding Kinetics of Riboswitches

Felix Kühnl

Department of Bioinformatics
University of Leipzig
felix@bioinf.uni-leipzig.de

Winterseminar 2016
Bled, Slovenia

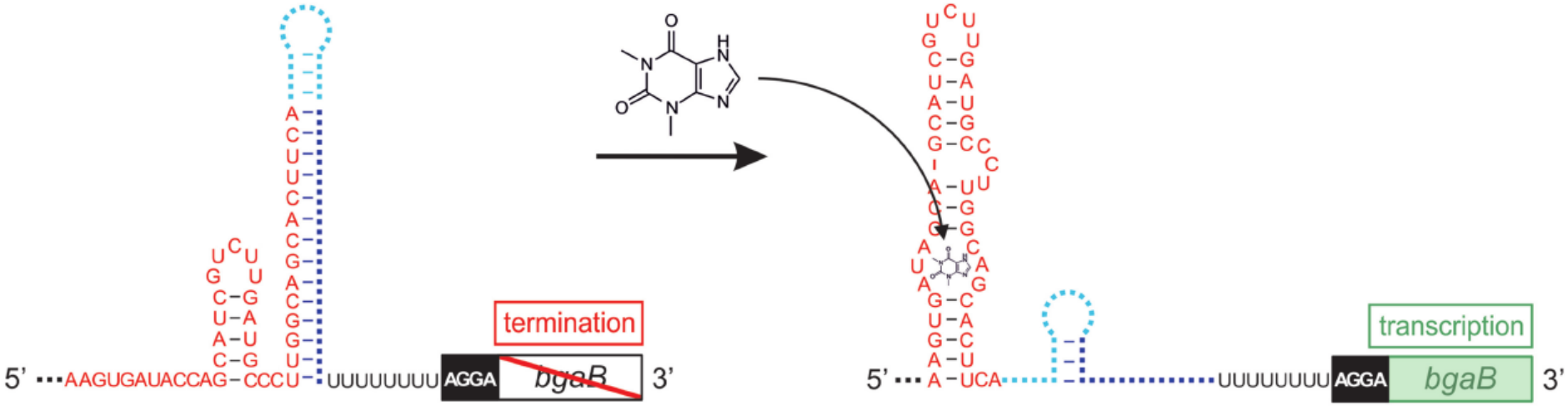
Riboswitches . . .

- . . . are regulatory RNA elements in 5′-UTR of genes
- regulate transcription or translation depending on a *ligand*
- consist of . . .
  - an *aptamer*: contains *binding pocket* for ligand L
  - an *actuator*, e. g. a *terminator hairpin*
- binding of ligand yields *energy bonus* $\theta_L < 0$

  $\Longrightarrow E_{L\varphi} = E_\varphi + \theta_L$
- time-critical interaction with RNA polymerase / ribosome

Riboswitches . . .

- . . . are regulatory RNA elements in 5′-UTR of genes
- regulate transcription or translation depending on a *ligand*

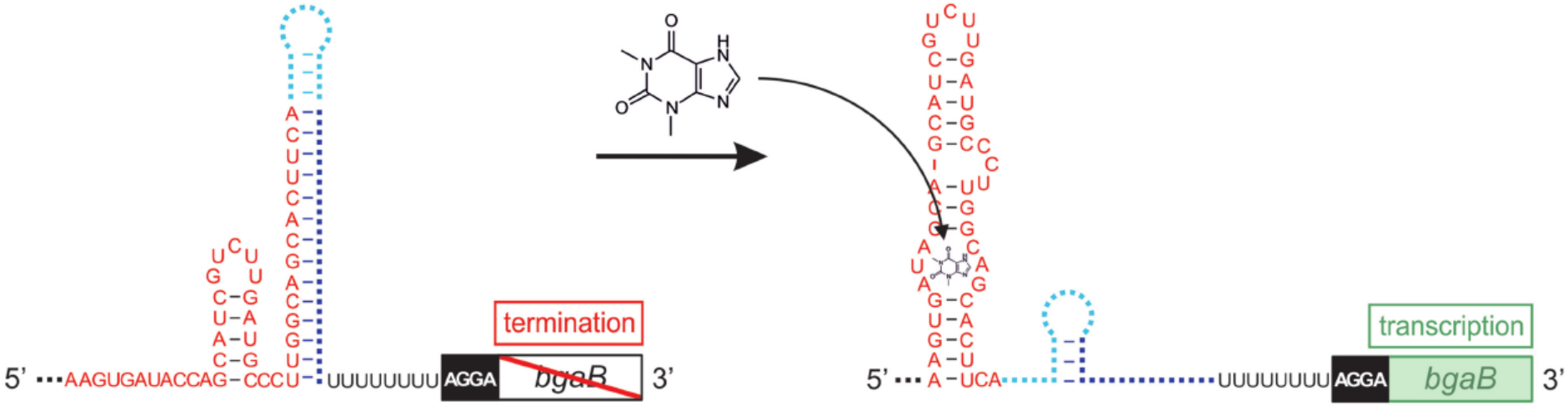


transcriptional, theophylline-controlled ON switch (figure: Wachsmuth et al. 2012)

- consist of . . .
  - an *aptamer*: contains *binding pocket* for ligand L
  - an *actuator*, e. g. a *terminator hairpin*
- binding of ligand yields *energy bonus* $\theta_L < 0$
  $\Longrightarrow E_{L\varphi} = E_\varphi + \theta_L$
- time-critical interaction with RNA polymerase / ribosome

Riboswitches . . .

- . . . are regulatory RNA elements in 5′-UTR of genes
- regulate transcription or translation depending on a *ligand*



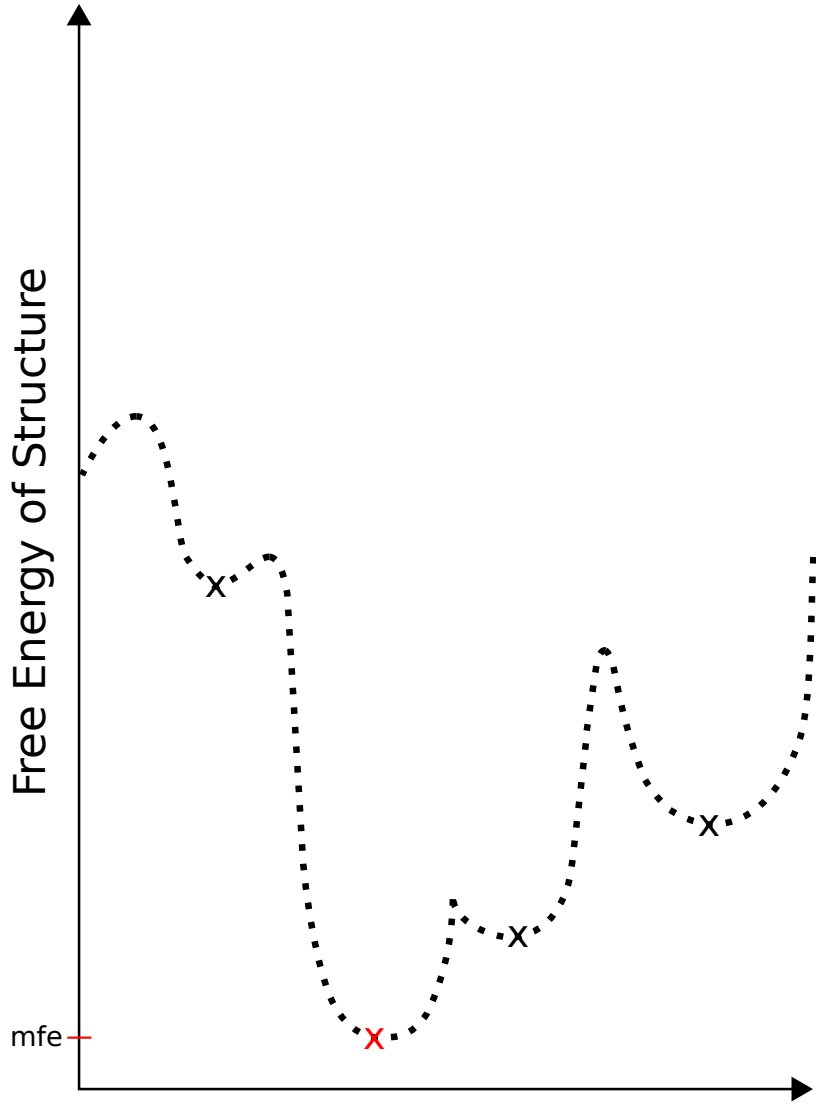
transcriptional, theophylline-controlled ON switch (figure: Wachsmuth et al. 2012)

- consist of . . .
  - an *aptamer*: contains *binding pocket* for ligand L
  - an *actuator*, e. g. a *terminator hairpin*
- binding of ligand yields *energy bonus* $\theta_L < 0$

  $\Longrightarrow E_{L\varphi} = E_\varphi + \theta_L$
- time-critical interaction with RNA polymerase / ribosome

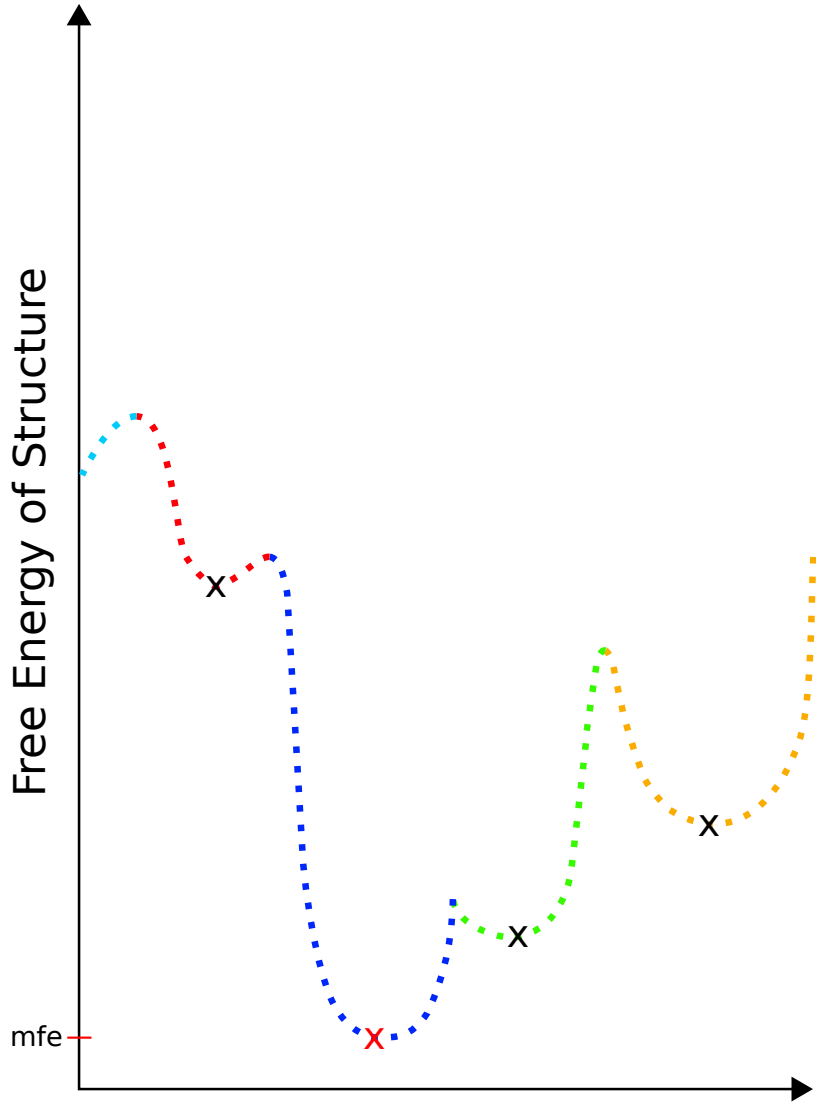
RNA folding kinetics: energy landscapes

- structure $\varphi \mapsto$ energy value (TURNER energy model)
- *transitions* between structures
- *gradient basins* as macrostates

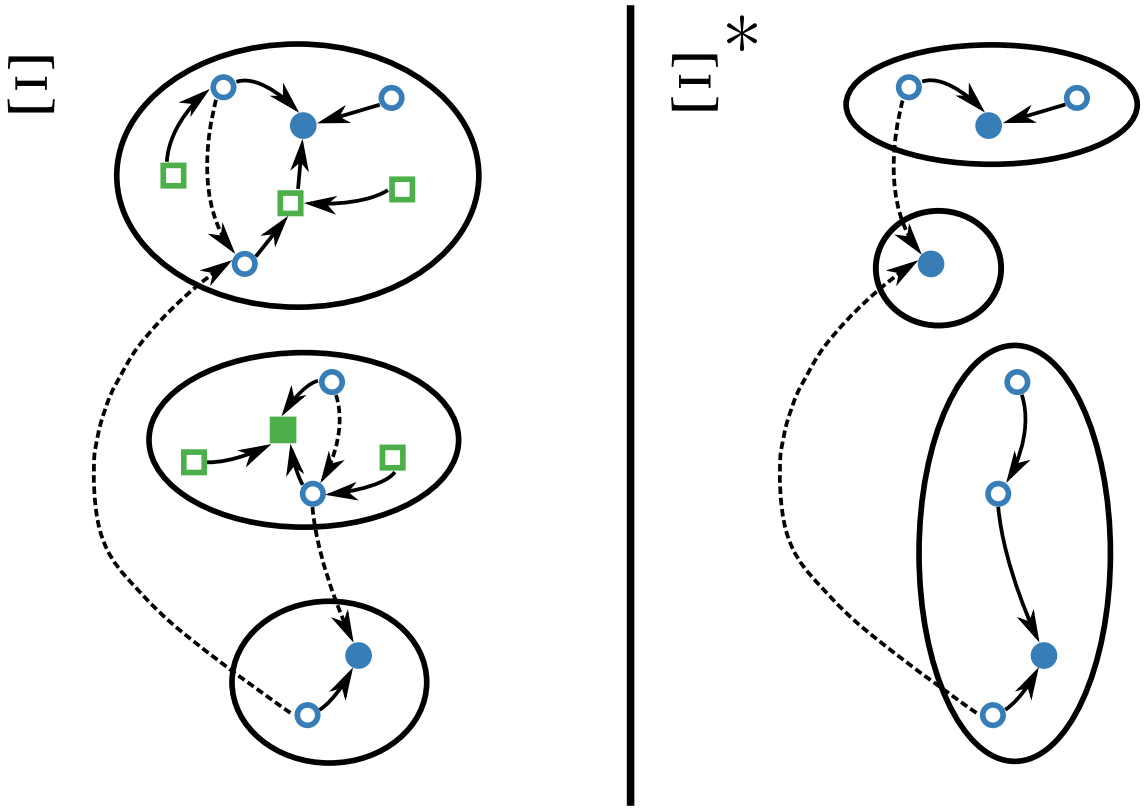


RNA folding kinetics: energy landscapes

- structure $\varphi \mapsto$ energy value (TURNER energy model)
- *transitions* between structures
- *gradient basins* as macrostates



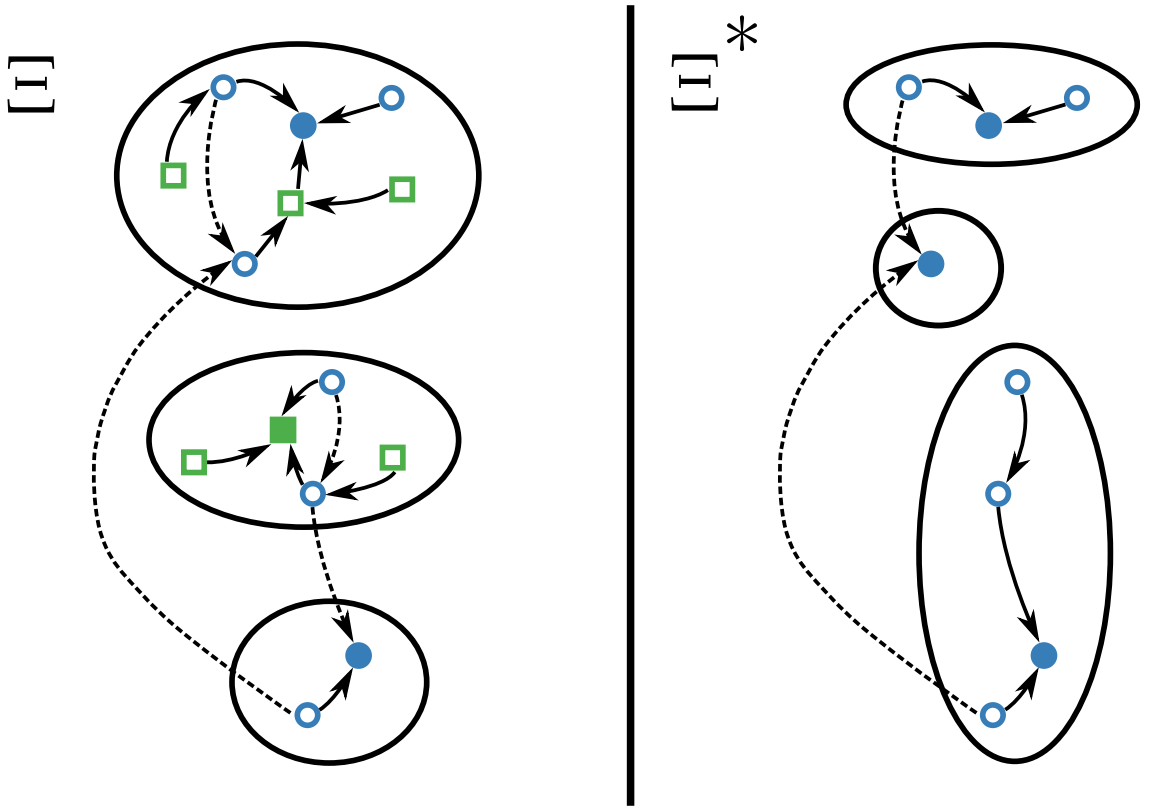
Riboswitch energy landscapes



$R_i \rightleftharpoons R_j$ $L + R_i \rightleftharpoons LR_j$ $LR_i \rightleftharpoons LR_j$

(R – RNA L – ligand LR – dimer)

Riboswitch energy landscapes



$R_i \rightleftharpoons R_j$ $\quad$ $L + R_i \rightleftharpoons LR_j$ $\quad$ $LR_i \rightleftharpoons LR_j$

(R – RNA $\quad$ L – ligand $\quad$ LR – dimer)

Modelling the system

by first-order rate laws: (assumption: $[L] = l_0 = \text{const}$)

$$(i = 1, \ldots, n) \qquad [\dot{R}_i] = + \sum_{\substack{1 \le k \le n \\ k \neq i}} r_{i \leftarrow k}[R_k] + \sum_{\substack{1 \le k \le m \\ k \neq i}} r_{i \leftarrow Lk}[LR_k] - \sum_{\substack{1 \le k \le n \\ k \neq i}} r_{k \leftarrow i}[R_i] - \sum_{\substack{1 \le k \le m \\ k \neq i}} r_{Lk \leftarrow i} \cdot l_0 \cdot [R_i],$$

$$(j = 1, \ldots, m) \qquad [\dot{LR}_j] = + \sum_{\substack{1 \le k \le n \\ k \neq j}} r_{Lj \leftarrow k} \cdot l_0 \cdot [R_k] + \sum_{\substack{1 \le k \le m \\ k \neq j}} r_{Lj \leftarrow Lk}[LR_k] - \sum_{\substack{1 \le k \le n \\ k \neq j}} r_{k \leftarrow Lj}[LR_j] - \sum_{\substack{1 \le k \le m \\ k \neq j}} r_{Lk \leftarrow Lj}[LR_j].$$

for n monomer and m dimer basins.

Modelling the system

by first-order rate laws: (assumption: $[L] = l_0 = \text{const}$)

$$(i = 1, \ldots, n) \qquad [\dot{R}_i] = + \sum_{\substack{1 \le k \le n \\ k \neq i}} r_{i \leftarrow k} [R_k] + \sum_{\substack{1 \le k \le m \\ k \neq i}} r_{i \leftarrow Lk} [LR_k] - \sum_{\substack{1 \le k \le n \\ k \neq i}} r_{k \leftarrow i} [R_i] - \sum_{\substack{1 \le k \le m \\ k \neq i}} r_{Lk \leftarrow i} \cdot l_0 \cdot [R_i],$$

$$(j = 1, \ldots, m) \qquad [\dot{LR}_j] = + \sum_{\substack{1 \le k \le n \\ k \neq j}} r_{Lj \leftarrow k} \cdot l_0 \cdot [R_k] + \sum_{\substack{1 \le k \le m \\ k \neq j}} r_{Lj \leftarrow Lk} [LR_k] - \sum_{\substack{1 \le k \le n \\ k \neq j}} r_{k \leftarrow Lj} [LR_j] - \sum_{\substack{1 \le k \le m \\ k \neq j}} r_{Lk \leftarrow Lj} [LR_j].$$

for n monomer and m dimer basins.

Modelling the system

by first-order rate laws: (assumption: $[L] = l_0 = \text{const}$)

$$(i = 1, \ldots, n) \qquad [\dot{R}_i] = + \sum_{\substack{1 \le k \le n \\ k \neq i}} r_{i \leftarrow k}[R_k] + \sum_{1 \le k \le m} r_{i \leftarrow Lk}[LR_k]$$
$$- \sum_{\substack{1 \le k \le n \\ k \neq i}} r_{k \leftarrow i}[R_i] - \sum_{1 \le k \le m} r_{Lk \leftarrow i} \cdot l_0 \cdot [R_i],$$

$$(j = 1, \ldots, m) \qquad [\dot{LR}_j] = + \sum_{\substack{1 \le k \le n \\ k \neq j}} r_{Lj \leftarrow k} \cdot l_0 \cdot [R_k] + \sum_{1 \le k \le m} r_{Lj \leftarrow Lk}[LR_k]$$
$$- \sum_{\substack{1 \le k \le n \\ k \neq j}} r_{k \leftarrow Lj}[LR_j] - \sum_{1 \le k \le m} r_{Lk \leftarrow Lj}[LR_j].$$

for n monomer and m dimer basins.

Modelling the system

by first-order rate laws: (assumption: $[L] = l_0 = \text{const}$)

$$(i = 1, \ldots, n) \qquad [\dot{R}_i] = + \sum_{\substack{1 \le k \le n \\ k \neq i}} r_{i \leftarrow k}[R_k] + \sum_{\substack{1 \le k \le m \\ k \neq i}} r_{i \leftarrow Lk}[LR_k]$$

$$- \sum_{\substack{1 \le k \le n \\ k \neq i}} r_{k \leftarrow i}[R_i] - \sum_{\substack{1 \le k \le m \\ k \neq i}} r_{Lk \leftarrow i} \cdot l_0 \cdot [R_i],$$

$$(j = 1, \ldots, m) \qquad [\dot{LR}_j] = + \sum_{\substack{1 \le k \le n \\ k \neq j}} r_{Lj \leftarrow k} \cdot l_0 \cdot [R_k] + \sum_{\substack{1 \le k \le m \\ k \neq j}} r_{Lj \leftarrow Lk}[LR_k]$$

$$- \sum_{\substack{1 \le k \le n \\ k \neq j}} r_{k \leftarrow Lj}[LR_j] - \sum_{\substack{1 \le k \le m \\ k \neq j}} r_{Lk \leftarrow Lj}[LR_j].$$

for n monomer and m dimer basins.

Modelling the system

by first-order rate laws: (assumption: $[L] = l_0 = \text{const}$)

$$(i = 1, \ldots, n) \qquad [\dot{R}_i] = + \sum_{\substack{1 \le k \le n \\ k \neq i}} r_{i \leftarrow k}[R_k] + \sum_{\substack{1 \le k \le m \\ k \neq i}} r_{i \leftarrow Lk}[LR_k]$$
$$- \sum_{\substack{1 \le k \le n \\ k \neq i}} r_{k \leftarrow i}[R_i] - \sum_{\substack{1 \le k \le m \\ k \neq i}} r_{Lk \leftarrow i} \cdot l_0 \cdot [R_i],$$

$$(j = 1, \ldots, m) \qquad [\dot{LR}_j] = + \sum_{\substack{1 \le k \le n \\ k \neq j}} r_{Lj \leftarrow k} \cdot l_0 \cdot [R_k] + \sum_{\substack{1 \le k \le m \\ k \neq j}} r_{Lj \leftarrow Lk}[LR_k]$$
$$- \sum_{\substack{1 \le k \le n \\ k \neq j}} r_{k \leftarrow Lj}[LR_j] - \sum_{\substack{1 \le k \le m \\ k \neq j}} r_{Lk \leftarrow Lj}[LR_j].$$

for n monomer and m dimer basins.

Modelling the system

by first-order rate laws: (assumption: $[L] = l_0 = \text{const}$)

$$(i = 1, \ldots, n) \qquad [\dot{R}_i] = + \sum_{\substack{1 \le k \le n \\ k \neq i}} r_{i \leftarrow k}[R_k] + \sum_{\substack{1 \le k \le m \\ k \neq i}} r_{i \leftarrow Lk}[LR_k] - \sum_{\substack{1 \le k \le n \\ k \neq i}} r_{k \leftarrow i}[R_i] - \sum_{\substack{1 \le k \le m \\ k \neq i}} r_{Lk \leftarrow i} \cdot l_0 \cdot [R_i],$$

$$(j = 1, \ldots, m) \qquad [\dot{LR}_j] = + \sum_{\substack{1 \le k \le n \\ k \neq j}} r_{Lj \leftarrow k} \cdot l_0 \cdot [R_k] + \sum_{\substack{1 \le k \le m \\ k \neq j}} r_{Lj \leftarrow Lk}[LR_k] - \sum_{\substack{1 \le k \le n \\ k \neq j}} r_{k \leftarrow Lj}[LR_j] - \sum_{\substack{1 \le k \le m \\ k \neq j}} r_{Lk \leftarrow Lj}[LR_j].$$

for n monomer and m dimer basins.

Simple, isn't it?

After some clever rewriting ...

$$\dot{x} = R(l_0)x$$

with ...

- $x = ([R_1], \ldots, [R_n], [LR_1], \ldots, [LR_m])^T$ – vector of concentrations
- $R(l_0)$ – rate coefficient matrix, more detailed:

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

where

- A – monomer rates, $(a_{ji}) = r_{j \leftarrow i}$
- B – dimer rates, $(b_{ji}) = r_{Lj \leftarrow Li}$
- C – dimer-monomer rates, $(c_{ji}) = r_{Lj \leftarrow i}$
- D – monomer-dimer rates, $(d_{ji}) = r_{j \leftarrow Li}$

Simple, isn't it?

After some clever rewriting ...

$$\dot{x} = R(l_0)x$$

with ...

- $x = ([R_1], \ldots, [R_n], [LR_1], \ldots, [LR_m])^T$ – vector of concentrations
- $R(l_0)$ – rate coefficient matrix, more detailed:

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

where

- A – monomer rates, $(a_{ji}) = r_{j \leftarrow i}$
- B – dimer rates, $(b_{ji}) = r_{Lj \leftarrow Li}$
- C – dimer-monomer rates, $(c_{ji}) = r_{Lj \leftarrow i}$
- D – monomer-dimer rates, $(d_{ji}) = r_{j \leftarrow Li}$

Rate coefficients

- general form (Arrhenius equation):

$$r_{\chi \leftarrow \varphi} = a_{\text{RNA}} \cdot \exp\left(-\frac{E_a(\chi \leftarrow \varphi)}{RT}\right)$$

  for RNA structures φ and χ
  (E_a – activation energy R – gas constant T – temperature)
- a_{RNA} – "collision rate" (from experiment)
- Activation energy (Metropolis rule):

$$E_a(\chi \leftarrow \varphi) = \begin{cases} \max(E_\varphi, E_\chi) - E_\varphi & \text{if } \varphi \text{ and } \chi \text{ are neighbored} \\ \infty & \text{else} \end{cases}$$

- macrorates for basins i, j (Wolfinger et al.):

$$r_{j \leftarrow i} = \sum_{\substack{\varphi \in i \\ \chi \in j}} \Pr[\varphi | i] r_{\chi \leftarrow \varphi}$$

Rate coefficients

- general form (Arrhenius equation):

$$r_{\chi \leftarrow \varphi} = a_{\text{RNA}} \cdot \exp\left(-\frac{E_a(\chi \leftarrow \varphi)}{RT}\right)$$

  for RNA structures φ and χ
  (E_a – activation energy R – gas constant T – temperature)
- a_{RNA} – "collision rate" (from experiment)
- Activation energy (Metropolis rule):

$$E_a(\chi \leftarrow \varphi) = \begin{cases} \max(E_\varphi, E_\chi) - E_\varphi & \text{if } \varphi \text{ and } \chi \text{ are neighbored} \\ \infty & \text{else} \end{cases}$$

- macrorates for basins i, j (Wolfinger et al.):

$$r_{j \leftarrow i} = \sum_{\substack{\varphi \in i \\ \chi \in j}} \Pr[\varphi | i] r_{\chi \leftarrow \varphi}$$

Rate coefficients

- general form (Arrhenius equation):

$$r_{\chi \leftarrow \varphi} = a_{\text{RNA}} \cdot \exp\left(-\frac{E_a(\chi \leftarrow \varphi)}{RT}\right)$$

  for RNA structures φ and χ
  (E_a – activation energy R – gas constant T – temperature)
- a_{RNA} – "collision rate" (from experiment)
- Activation energy (Metropolis rule):

$$E_a(\chi \leftarrow \varphi) = \begin{cases} \max(E_\varphi, E_\chi) - E_\varphi & \text{if } \varphi \text{ and } \chi \text{ are neighbored} \\ \infty & \text{else} \end{cases}$$

- macrorates for basins i, j (Wolfinger et al.):

$$r_{j \leftarrow i} = \sum_{\substack{\varphi \in i \\ \chi \in j}} \Pr[\varphi | i] r_{\chi \leftarrow \varphi}$$

Rate coefficients ctd.

Lemma

1. *Rate for $LR_i \to LR_j$ (dimer-dimer rate):*

$$r_{L\chi \leftarrow L\varphi} = r_{\chi \leftarrow \varphi} \qquad r_{Lj \leftarrow Li} = r_{j \leftarrow i}$$

2. *Rate for $L + R_i \to LR_j$ (dimerization rate):*

$$r_{L\varphi \leftarrow \varphi} = a_{dim} \qquad r_{Lj \leftarrow i} = a_{dim} \cdot \frac{Z[i \cap j]}{Z[i]}$$

3. *Rate for $LR_i \to L + R_j$ (dissociation rate):*

$$r_{\varphi \leftarrow L\varphi} = a_{dim} \cdot \exp(\theta_L / RT) \qquad r_{j \leftarrow Li} = a_{dim} \cdot \frac{Z[i \cap j]}{Z[i]} \cdot \exp(\theta_L / RT)$$

$\theta_L < 0$ – *ligand binding bonus*

Rate coefficients ctd.

Lemma

1. *Rate for $LR_i \rightarrow LR_j$ (dimer-dimer rate):*

$$r_{L\chi \leftarrow L\varphi} = r_{\chi \leftarrow \varphi} \qquad r_{Lj \leftarrow Li} = r_{j \leftarrow i}$$

2. *Rate for $L + R_i \rightarrow LR_j$ (dimerization rate):*

$$r_{L\varphi \leftarrow \varphi} = a_{dim} \qquad r_{Lj \leftarrow i} = a_{dim} \cdot \frac{Z[i \cap j]}{Z[i]}$$

3. *Rate for $LR_i \rightarrow L + R_j$ (dissociation rate):*

$$r_{\varphi \leftarrow L\varphi} = a_{dim} \cdot \exp(\theta_L / RT) \qquad r_{j \leftarrow Li} = a_{dim} \cdot \frac{Z[i \cap j]}{Z[i]} \cdot \exp(\theta_L / RT)$$

$\theta_L < 0$ – ligand binding bonus

Rate coefficients ctd.

Lemma

1. *Rate for* $LR_i \rightarrow LR_j$ *(dimer-dimer rate):*

$$r_{L\chi \leftarrow L\varphi} = r_{\chi \leftarrow \varphi} \qquad r_{Lj \leftarrow Li} = r_{j \leftarrow i}$$

2. *Rate for* $L + R_i \rightarrow LR_j$ *(dimerization rate):*

$$r_{L\varphi \leftarrow \varphi} = a_{dim} \qquad r_{Lj \leftarrow i} = a_{dim} \cdot \frac{Z[i \cap j]}{Z[i]}$$

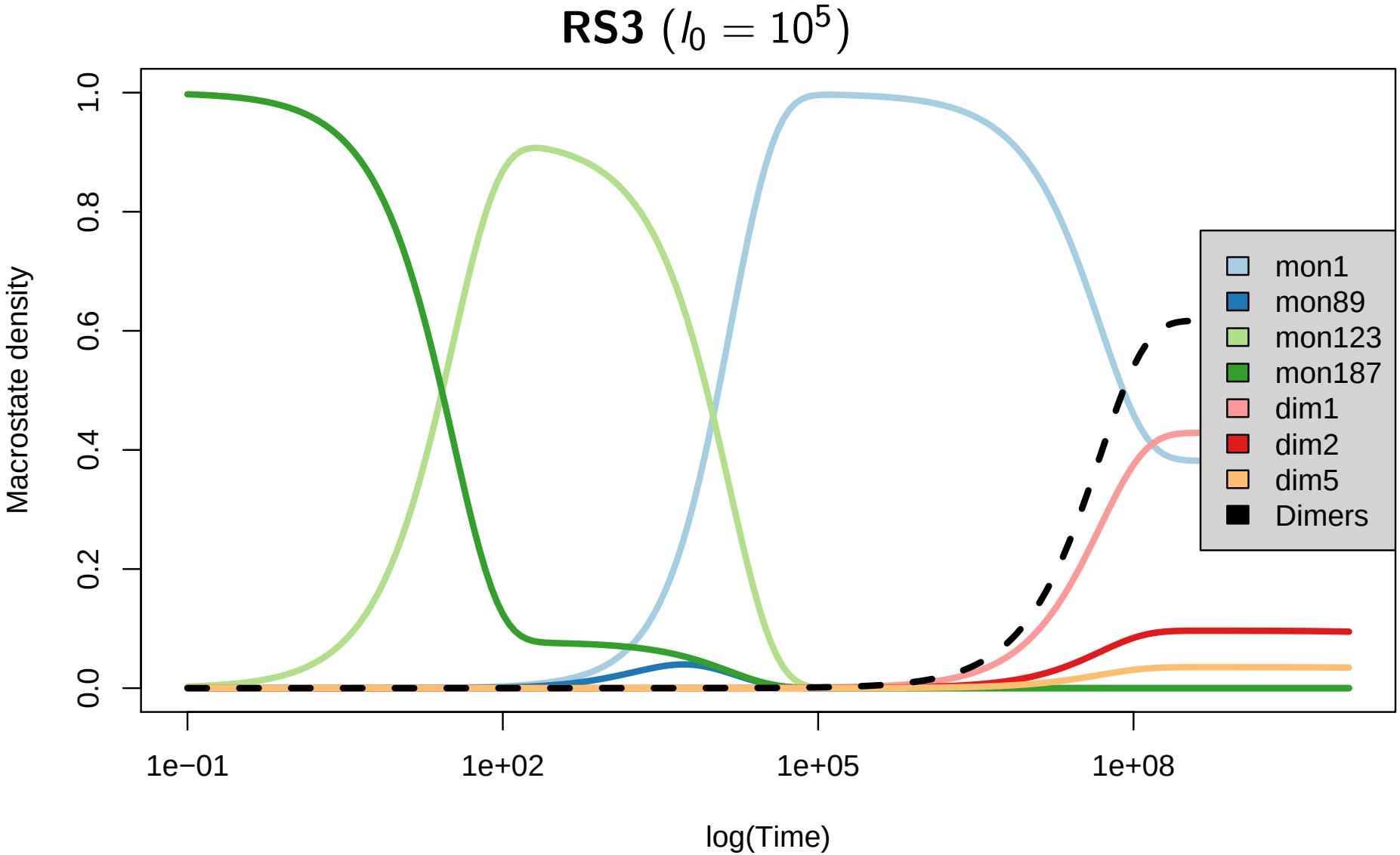
3. *Rate for* $LR_i \rightarrow L + R_j$ *(dissociation rate):*

$$r_{\varphi \leftarrow L\varphi} = a_{dim} \cdot \exp(\theta_L / RT) \qquad r_{j \leftarrow Li} = a_{dim} \cdot \frac{Z[i \cap j]}{Z[i]} \cdot \exp(\theta_L / RT)$$

$\theta_L < 0$ – *ligand binding bonus*

Results

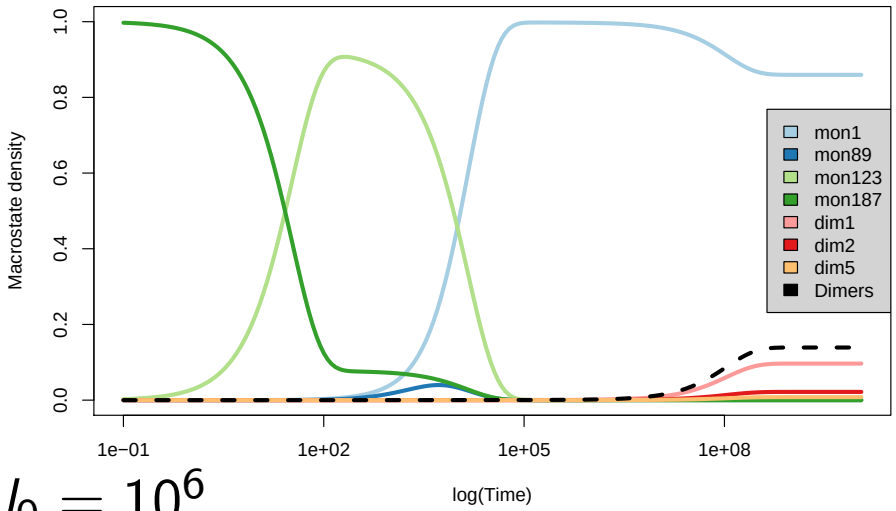
`RNAsubopt` → `barriers` + "monomer↔dimer" rates → `treekin`



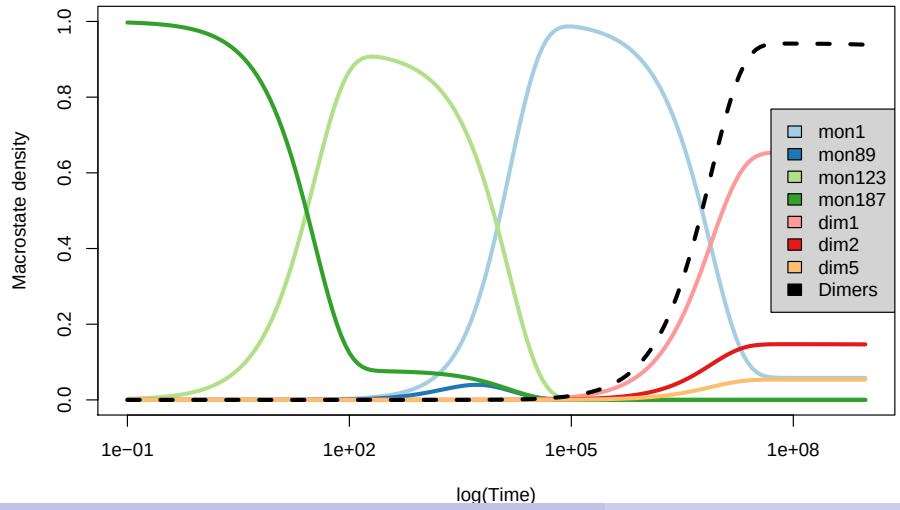
Results ctd.

RS3

$I_0 = 10^4$

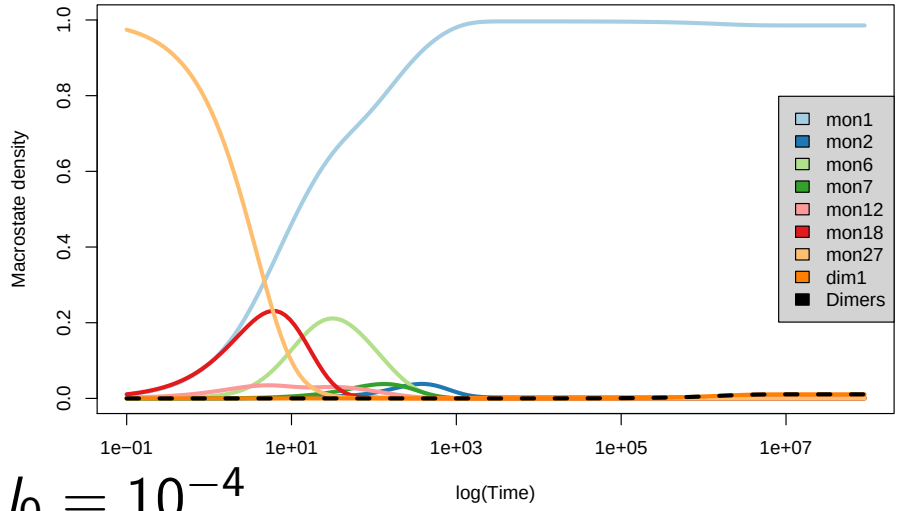


$I_0 = 10^6$

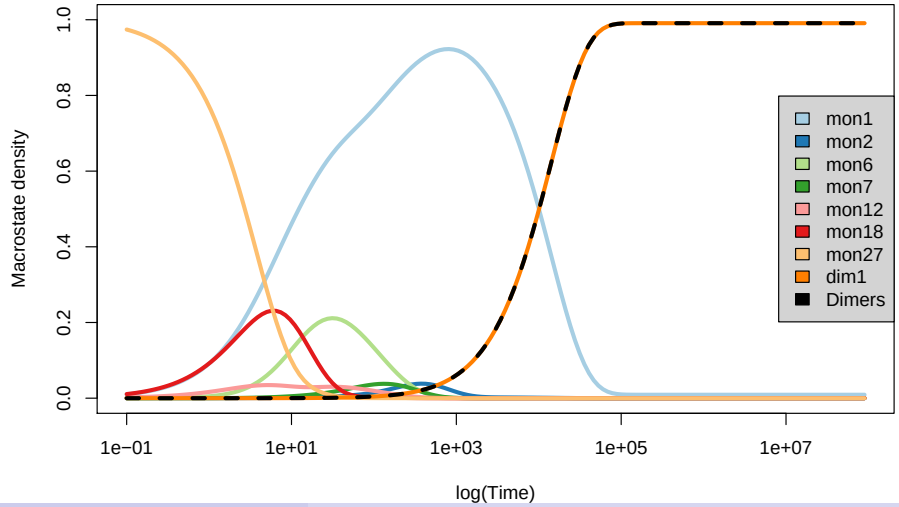


RS3 shortened

$I_0 = 10^{-8}$



$I_0 = 10^{-4}$



Outlook

- goal: verify functionality of artificial riboswitch designs
- estimates of collision rates for RNA folding and dimerization:

$$r = (a_{\text{RNA}} \mid a_{\text{dim}}) \cdot \exp\left(\frac{E_a}{RT}\right)$$

  - a_{dim}: different for each aptamer and ligand
  - a_{RNA}: closing first base pair of loop *vs.* last base pair of stem