Co-transcriptional Folding Kinetics and Riboswitch Modeling

Ivo L. Hofacker

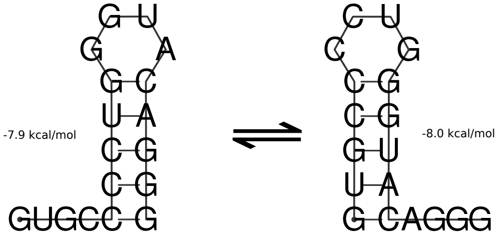
Institute for Theoretical Chemistry
Research Group Bioinformatics and Computational Biology
University of Vienna

Benasque RNA 2018

Thermodynamic vs. Kinetic Folding

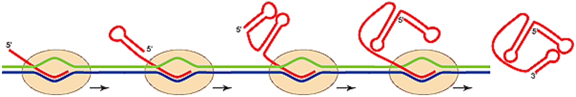
Equilibrium properties for RNA secondary strutcures can be calculated efficiently
But what about dynamics?

- On what time scale is equilibrium reached?
- How fast/slow is re-folding between dissimilar structures?
- What structures are populated initially?



Folding during Transcription

Almost all RNA structures may be affected by co-transcriptional folding:



- RNA is transcribed at a rate of 25–50 nucleotides per second
- Nascent chain starts folding as soon as its leaves the ribosome
- Helices once formed may be too stable to refold later on
- Co-transcriptional folding may drive the folding process to a well-defined folded state (possibly different from the MFE)
- An energy barrier of 5kcal/mol can prevent refolding during extension

Folding Dynamics as Markov Process

Let's compute prob. $P_x(t)$ of observing structure x at time t. Given transition rates k_{xy}, this gives rise to a *Markov process* with master equation

$$\frac{dP_x(t)}{dt} = \sum_{y \neq x} [P_y(t) k_{xy} - P_x(t) k_{yx}].$$

or in matrix form, with $k_{xx} = -\sum_{x \neq y} k_{yx}$:

$$\frac{d}{dt} P(t) = \mathbf{K} P(t).$$

A *formal* solution can be written simply

$$P(t) = e^{t \cdot \mathbf{K}} P(0)$$

Way too many states to solve directly (10^{17} for a tRNA)

Three Strategies for Predicting Folding Kinetics

- Folding trajectories via Monte-Carlo simulation
  - Time-consuming
  - Need statistics over many trajectories
  - Non-trivial to analyze and interpret
  - `kinfold`, `KineFold`
- Coarse grained dynamics via Barriers / Treekin / Barmap
  - Identify local minima, assign macro-states
  - Energy barriers and transition rates (`barriers`)
  - Solve $P_x(t)$ on coarse grained landscape (`treeekin`)
  - Extend sequence and transfer population to next landscape (`barmap`)
- Heuristic landscape construction
  - Model landscape by small set of representative structures
  - Estimate energy barriers and rates
  - Can be nicely combined with co-transcriptional folding `DrTransformer`

Stochastic Simulations

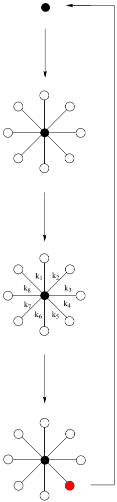
Simulate folding kinetics by Gillespie (rejectionless Monte Carlo) algorithm :

Generate all neighbors using a move-set
Close base pair – Open base pair

Assign rates to each move, e.g.:

$$k_i = \Gamma \cdot \min\left\{1, \exp\left(-\frac{\Delta E}{kT}\right)\right\}$$

Select a move i with probability $\propto k_i$
Advance clock by $1/\sum_i k_i$ (on average).

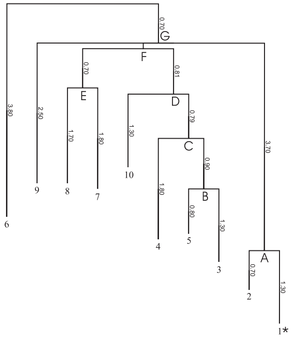


- ☹ computationally somewhat expensive
- ☹ need to analyze many trajectories
- ☺ easy to include co-transcriptional folding

RNA Landscape Analysis

Barrier trees

- Contains all local minima as leafs
- Barrier heights and saddles between minima
- Groups structures into *macro states*
- Transition rates between macro states
  → coarse grained dynamics
- Time and space proportional to the size of the landscape
  Limited to RNA $<$ 100nt
- Sampling based heuristics for longer RNAs



Flamm et al, Z. Phys. Chem. (2002)

Calculating barrier trees

The flooding algorithm:

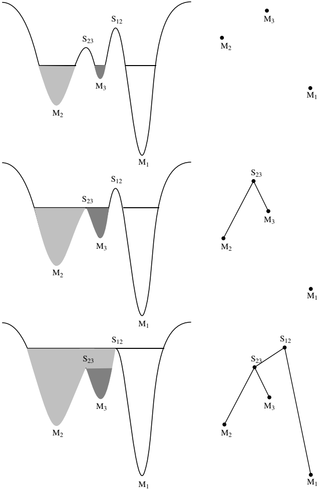
Read conformations in energy sorted order.

For each confirmation x we have three cases:

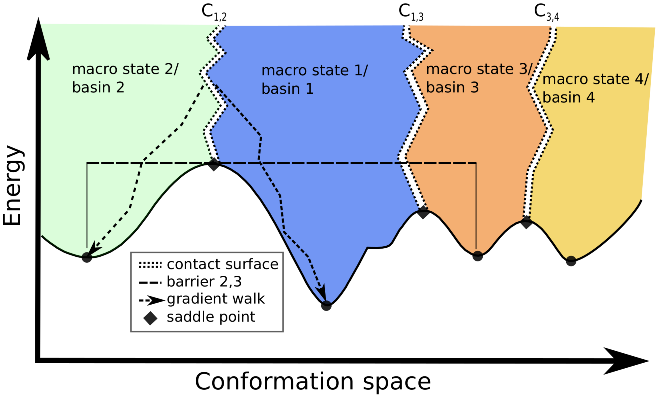
(a) x is a *local minimum* if it has no neighbors we've already seen

(b) x belongs to basin $B(s)$, if all known neighbors belong to $B(s)$

(c) if x has neighbors in several basins $B(s_1) \dots B(s_k)$ then it's a *saddle point* that *merges* these basins.



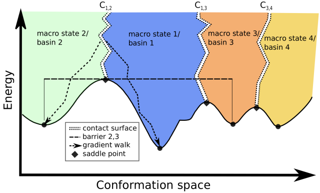
Coarse Graining the Landscape



Coarse Graining the folding dynamics

For a reduced description we need

- macro-states that form a partition of full configuration space
- transition rates between macro states
- macro-states defined via gradient walks



Transition rates could follow an Arrhenius rule $r_{\beta\alpha} = \exp\left(-(E^*_{\beta\alpha} - G_\alpha)/RT\right)$.

Better: include *all* transition states

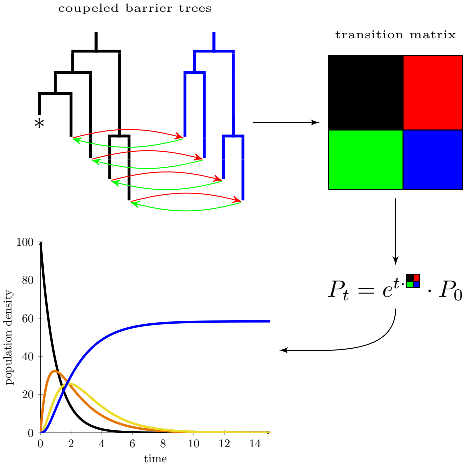
$$r_{\beta\alpha} = \sum_{y\in\beta}\sum_{x\in\alpha} r_{yx}\mathrm{Prob}[x|\alpha] \approx \frac{1}{Z_\alpha}\sum_{y\in\beta}\sum_{x\in\alpha} r_{yx}e^{-E(x)/RT}$$

assuming local equilibrium.

How to include Ligand Binding ?

- Need to know binding motif and binding rates from experiment
- Simple strategy:
  - Add binding energy $\theta = RT \ln \frac{K_d}{c^{\ominus}}$ to every binding competent structure
  - Assumes infinite ligand concentration and infinitely fast binding
- Treat binding / unbinding events explicitly
  - Barrier trees for bound and unbound states
  - Usual rates within bound / unbound structures
  - Concentration dependent rate of complex formation
    $k_{\text{off}} = k_{\text{on}} e^{-\theta/RT}, \quad r = k_{\text{on}} \cdot C$

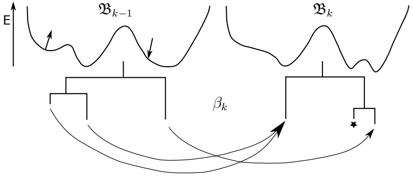
How to include Ligand Binding ?



Kühnl et al, BMC Bioinf. (2017), Wolfinger et al. Methods (2018)

Co-transcriptional folding with BarMap

Each extension of the RNA structure modifies the landscape:

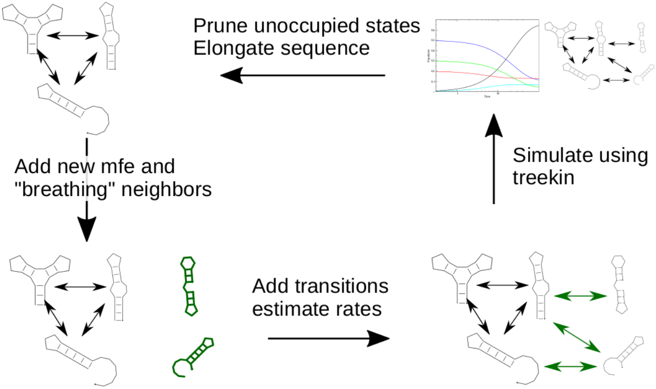


- Compute barrier trees for each sequence length $1 \dots n$
- Compute a mapping between the minima of subsequent landscapes
- Compute dynamics piece-wise:
  - Compute dynamics on landscape for length k
  - Transfer population to landscape of length $k+1$

Hofacker et al., RNA (2010)

DrTransformer: Fast co-transcriptional Folding

- Simulate a **small** network consisting only of the most relevant structural states
- Evolve network as RNA grows



Prune unoccupied states
Elongate sequence

Add new mfe and
"breathing" neighbors

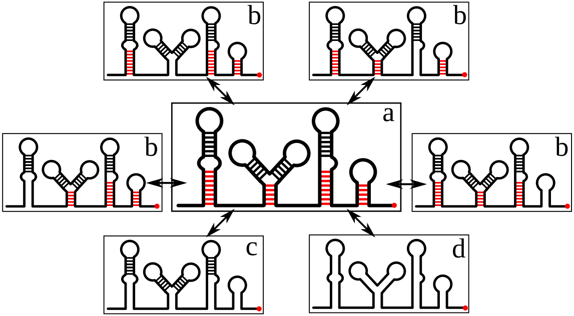
Add transitions
estimate rates

Simulate using
treekin

DrTransformer: “Breathing” neighbors

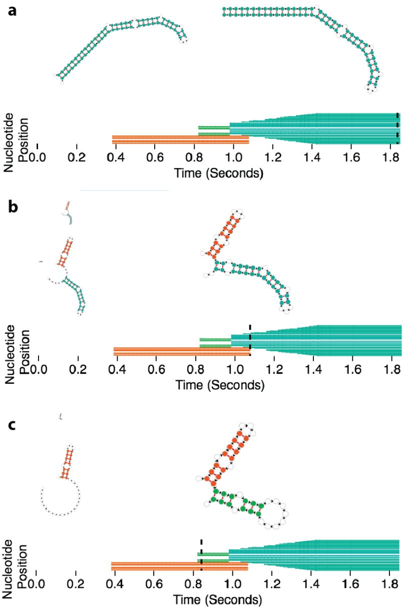
Which new structures should be added after an elongation step?

- Elongation can only effect the surroundings of the exterior loop
- Partially unfold all helices that protrude from exterior loop
- Use constrained folding to re-fold exterior loop surroundings



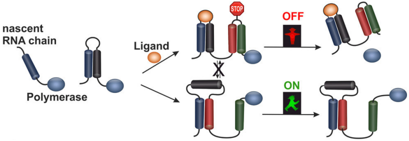
DrTransformer Visualization

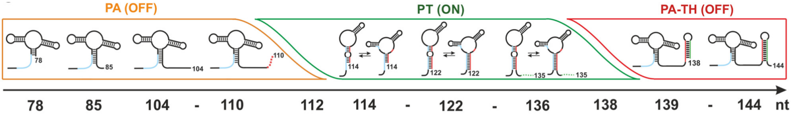
- Simple webinterface
- Interactive visualization Javascript and SVG
- Structure ensemble as function of time



Example: The dG-Riboswitch

- Aptamer for 2’deoxyguanosin
- Binding leads to transcription termination
- NMR analysis (Schwalbe lab): Ground state structure contains terminator even without ligand

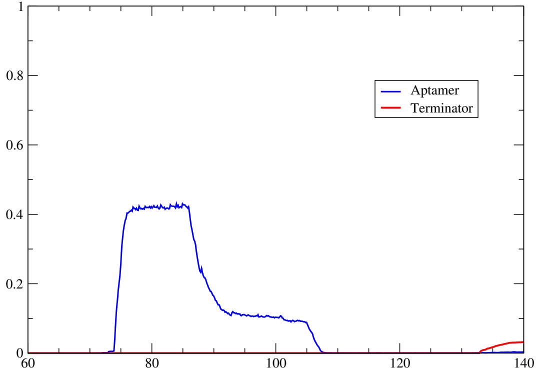




Helmling et al, JACS (2017)

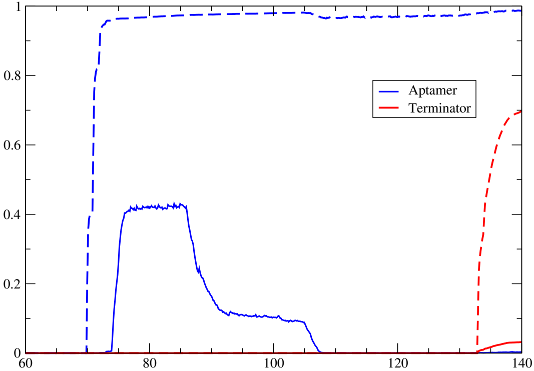
Kinfold simulation of the dG Riboswitch

- 10000 Kinfold trajectories (186 cpu hours)
- Classify each structure as aptamer and/or terminator
- Simulation with ligand: Add a bonus of 8kcal/mol for each binding competent structure



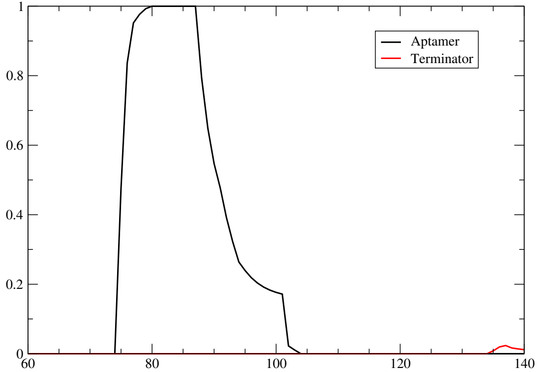
Kinfold simulation of the dG Riboswitch

- 10000 Kinfold trajectories (186 cpu hours)
- Classify each structure as aptamer and/or terminator
- Simulation with ligand: Add a bonus of 8kcal/mol for each binding competent structure

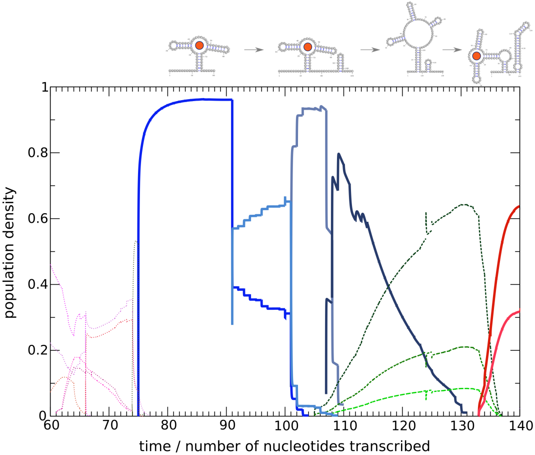


DrTrafo simulation of the dG Riboswitch

- Only 1 run needed (3 cpu sec)
- Classify each structure as aptamer and/or terminator
- Final state 1% population in terminator
- Simulation with ligand not yet possible



BarMap simulation of the dG Riboswitch



Simulation at 25C, transcription speed 25 nt/sec, ligand concentration of 1mM

Take home messages

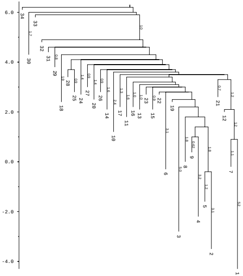
- RNAs don't always reach their MFE or equilibrium state in reasonable time.
- Co-transcriptional folding essential to regulatory elements such as riboswitches
- Predicting kinetics is much harder than predicting equilibrium
- Previous methods too slow too cumbersome
- Faster, easy to interpret methods, now available

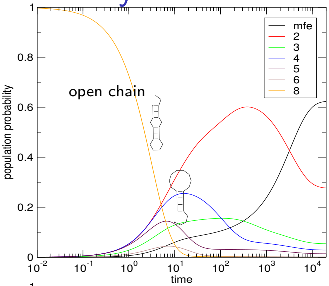
Acknowledgments

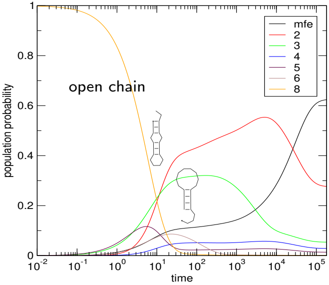
Christoph Flamm
Peter Stadler (Leipzig)
Ronny Lorenz (ViennaRNA)
Sebastian Will, Felix Kühnl (Leipzig, ligand modelling)
Michael **Wolfinger** (`barriers`, `treekin`)
Marcel **Kucharik**, Jing Qin (now SDU Odense) (`BGH`)
Stefan **Badelt** (now CalTech, `DrTransformer`)
Peter Kerpedjiev (now Harvard, DrTransformer visualization)
Stefan **Hammer** (visualization)

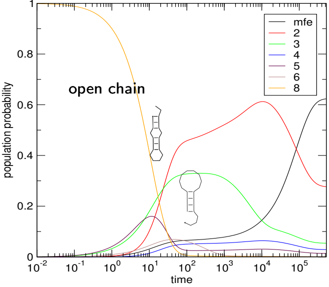
H. Schwalbe, B. Fürtig, Ch. Helmling (Frankfurt, (dG-ribsowitch)

Coarse grained dynamics vs. full dynamics



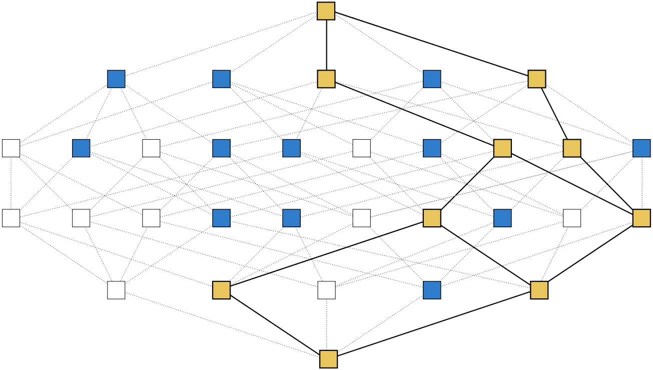






The findpath re-folding path heuristic

Perform a bounded breadth first search of direct paths.



- Only consider **direct** paths, i.e. where distance decreases with each step.
- Up to $D(x, y)!$ direct paths.
- Bound the search by keeping only *m* best candidates from each distance class.