Coarse grained RNA folding kinetics

Ronny Lorenz[1], Christoph Flamm[1], Ivo L. Hofacker[1]

[1]Institute for Theoretical Chemistry, University of Vienna, Waehringerstrasse 17, A-1090 Wien, Austria

GEN-AU GENOMFORSCHUNG IN ÖSTERREICH

Bioinformatics Integration Network

BMWF

Introduction

In recent years a growing number of RNA *switches*, i.e. regulatory RNAs whose function is coupled to a change in conformation, has been identified. The most prominent examples are RNA thermometers and riboswitches. This emphasizes the need for reliable and efficient methods to predict RNA folding kinetics.

While almost any equilibrium property on the level of RNA secondary structures can be computed by dynamic programming algorithms only a handfull of approaches exist for the prediction of RNA folding dynamics. For the most part these rely on stochastic folding simulations which require the time consuming sampling of a large number of trajectories.

In general the folding process can be defined by three key ingredients: (a) a state space of allowed conformations, (b) a move set that defines the elementary transitions between the states and (c) transition rates for each of the allowed transitions.

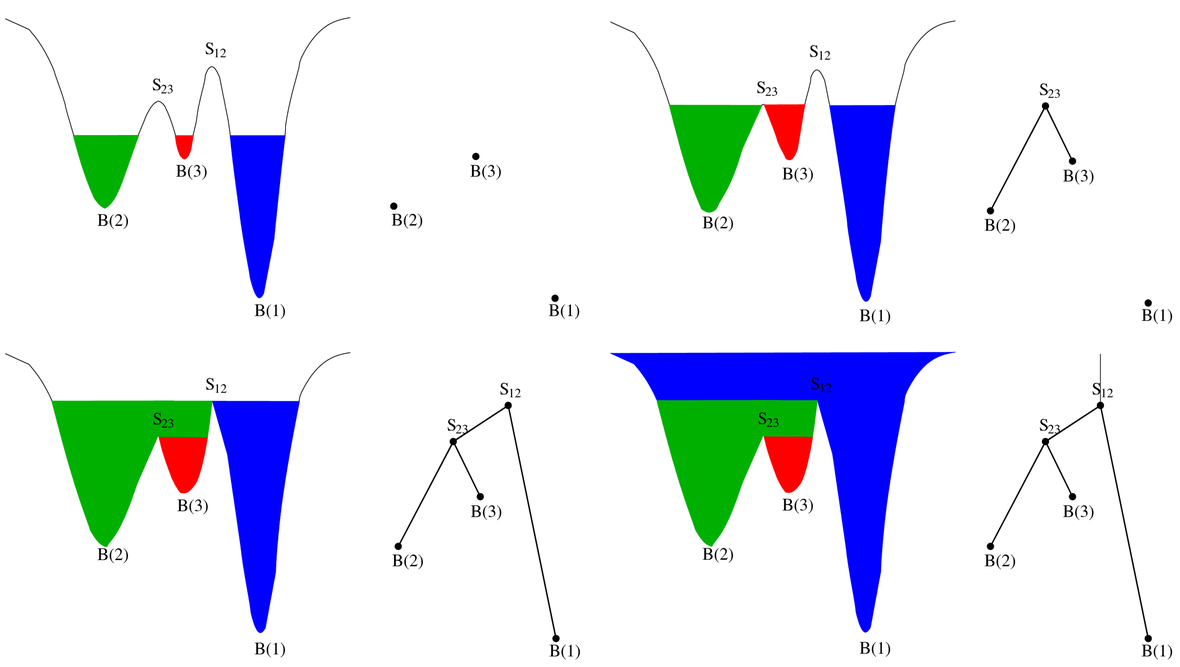
For RNA secondary structures the conformation space is the set of structures compatible with a given RNA sequence, the simplest move set consists of opening/closing an individual base pair and the Metropolis rule can be used for the transition rates.

With these ingredients RNA folding can be seen as a Markov process. Given an initial population distribution vector $\vec{p}(0)$ and the transition rate matrix $K=(r_{xy})$ the population densities of all states $\vec{p}(t)$ at any time t can be obtained from the master equation

$$\frac{d}{dt}\vec{p}(t)=K\cdot\vec{p}(t) \quad \text{with formal solution} \quad \vec{p}(t)=\mathrm{e}^{t\cdot K}\cdot\vec{p}(0) \tag{1}$$

The direct computation of the formal solution is infeasible except for trivially small examples. However, if coarse graining can reduce the number of conformations to a few thousand macro states, then the coarse grained dynamics can be computed efficiently by diagonalization of the rate matrix. This type of approach entails two tasks, (i) how to partition the conformation space into macro states, and (ii) how to compute effective transition rates between macro states. Here, we compare two possibilities for the partitioning, namely via gradient basins and via distance classes.

Barrier trees and gradient basins



The flooding algorithm *[1]* implemented in the barriers program processes an energy sorted list of all suboptimal structure states, in order to identify all local minima $B(x)$ and their connecting minimal saddle points S_{xy}. The result is a compact visualization of the energy landscape called barrier tree. Simultaneously, all structures x are assigned to their respective gradient basins (i.e. the local minimum that would be reached by a gradient walk starting from x). The gradient basins form a partition of the conformation space that provides a convenient coarse graining into macro-states. Transition rates between macro-states are estimated from the rates between micro-states *[2]*:

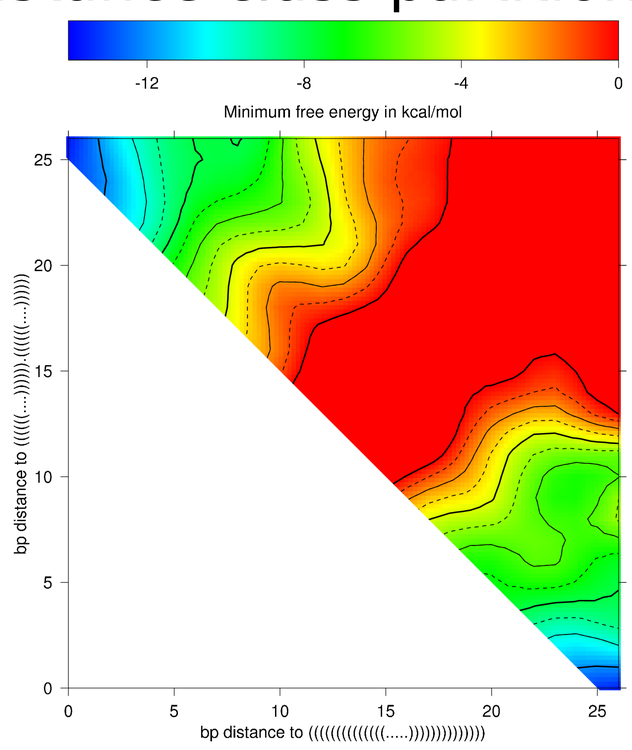
$$r_{\beta\alpha}=\sum_{x\in\alpha}\sum_{y\in\beta}k_{yx}\cdot\mathrm{Prob}[x|\alpha]\approx\sum_{x\in\alpha}\sum_{y\in\beta}k_{yx}\cdot\frac{\mathrm{e}^{\frac{-E(x)}{kT}}}{Q_\alpha} \tag{2}$$

The approximation assumes that equilibriation within the macro state is fast and thus the probability $\mathrm{Prob}[x|\alpha]$ equals its equilibrium value $\exp(-E(x)/kT)/Q_\alpha$. For the micro rates k_{xy} we usually use the Metropolis rule,

$$k_{xy}=\begin{cases} e^{-\frac{E(y)-E(x)}{kT}} & \text{if } E(x)<E(y) \\ 1 & \text{otherwise.}\end{cases}$$

While the gradient basin partitioning has been shown to yield results very similar to the exact dynamics, it relies on exhaustive enumeration of low-energy structures, and is therefore limited to RNA molecules no longer than some 100 nt.

Distance class partitioning



As a new alternative we present here a partitioning into distance classes with respect to two reference structures. Typically, we choose the ground state and one metastable state as reference structures. Minimum free energy structures and partition function for each class can be computed in $O(n^7)$ time using the *RNA2Dfold* program *[3]*. This results in a 2D projection of the energy landscape that is convenient for visualization.

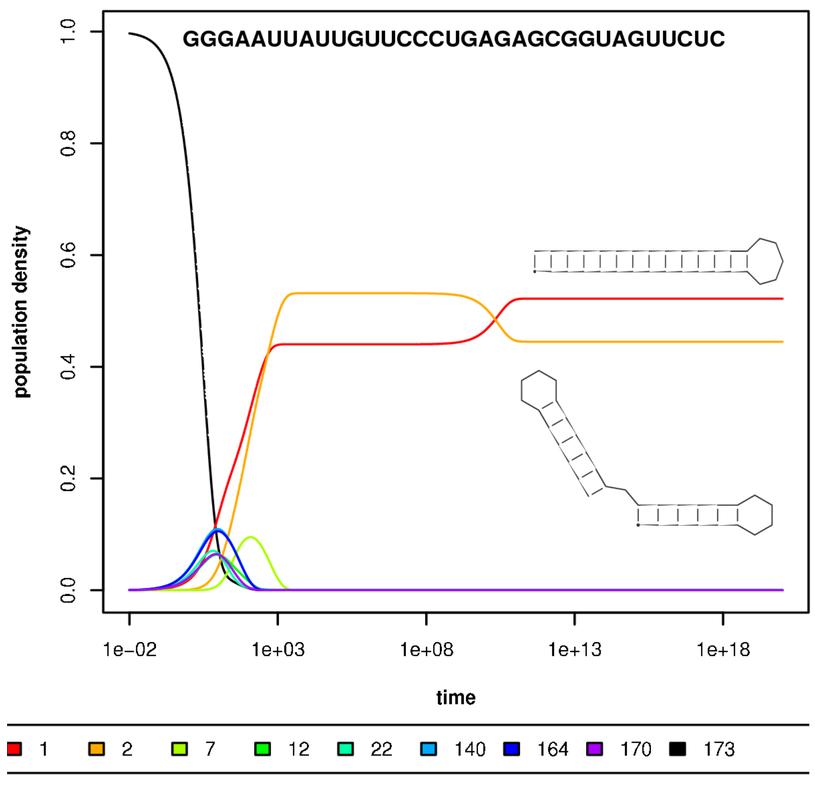
Transition rates are computed according to eq.(2), however, rather than enumerating all structures, we approximate the sum by Boltzmann sampling from each distance class. For each sample of Boltzmann distributed structures S_α we compute

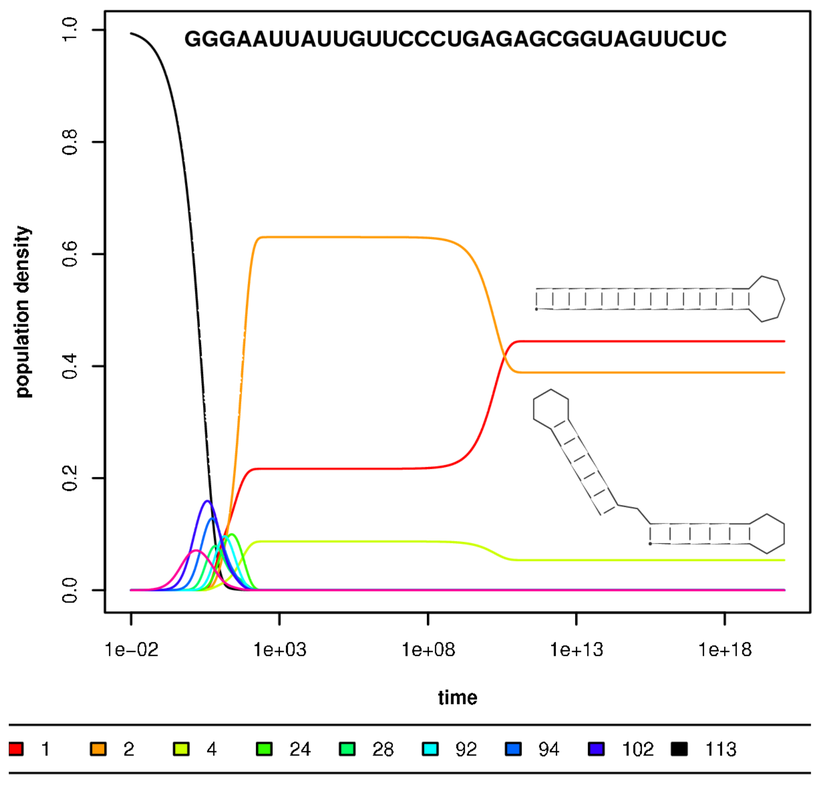
$$r_{\beta\alpha}\approx\frac{1}{|S_\alpha|}\sum_{x\in S_\alpha}\sum_{y\in\beta\cap N(x)}k_{yx} \tag{3}$$

Special care is taken that detailed balance is not effected by sampling errors. A sample size of 1000 per macro state proved sufficient for the examples tested.

As shown below, both methods yield qualitatively similar results. In the figure identical colors are used for macro states containing the same minimum. Since the rest of the structures will in general be different, exact occupancies and even equilibrium values differ between both methods.

Since *RNA2Dfold* works well for sequences up to 500 nt with this new approach for coarse grained kinetic folding, significantly longer RNAs can be used than in the barrier tree approach.





Acknowledgements

This work was supported by the Austrian Ministry of Science and Research GEN-AU project “Bioinformatics Integration Network II & III” and “noncoding RNA II” and the University of Vienna.

References

[1] Christoph Flamm, Ivo L. Hofacker, Peter F. Stadler, and Michael T. Wolfinger. Barrier trees of degenerate landscapes. *Z. Phys. Chem.*, 216:155–173, 2002.

[2] Michael T. Wolfinger, Andreas W. Svrcek-Seiler, Christoph Flamm, Ivo L. Hofacker, and Peter F. Stadler. Efficient computation of RNA folding dynamics. J. *Phys. A: Math. Gen.*, 37:4731–4741, 2004.

[3] Ronny Lorenz, Christoph Flamm, and Ivo L. Hofacker. 2D projections of RNA folding landscapes. In Ivo Grosse, Steffen Neumann, Stefan Posch, Falk Schreiber, and Peter F. Stadler, editors, *German Conference on Bioinformatics 2009*, volume 157 of *Lecture Notes in Informatics*, pages 11–20, Bonn, September 2009. Gesellschaft f. Informatik.