

RNA Secondary Structure Thermodynamics and Kinetics

Ronny Lorenz

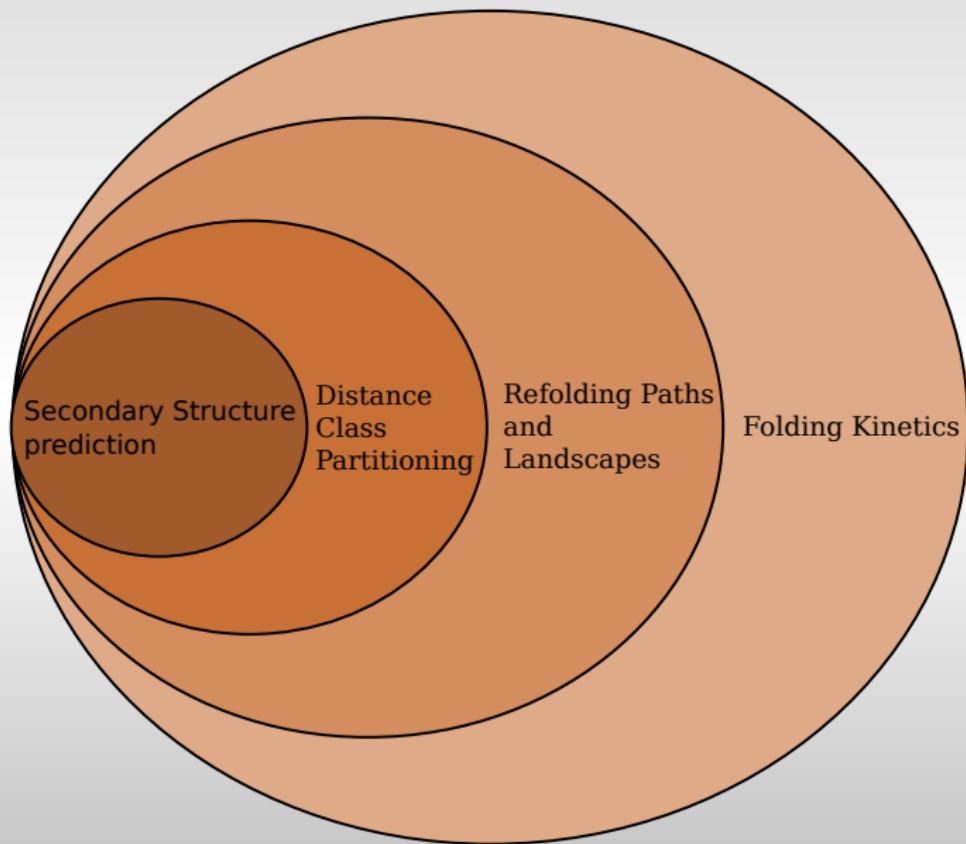
Institute for Theoretical Chemistry
University of Vienna

Vienna, Austria, October 5, 2014



A Journey to Vienna with RNA Structures and Schnitzel

Topics/Work of my PhD study



Topics/Work of my PhD study

Published work

- ① ViennaRNA Package 2.0¹
- ② RNA/DNA hybrid structure prediction²
- ③ Secondary structure prediction with G-Quadruplexes³
- ④ 2D projections of secondary structure landscapes⁴

Unpublished work

- ① Detection of alternative low-free energy secondary structure states
- ② Prediction of (near) optimal, indirect RNA refolding paths
- ③ Coarse grained RNA folding kinetics with *ab-initio* partitioning
- ④ Simulation of cotranscriptional folding dynamics

¹Lorenz et al., *Algorithms for Molecular Biology* 2011, 6:26

²Lorenz et al., *Bioinformatics* 2012 Vol. 28 no. 19

³Lorenz et al., *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2013 Volume 10, Issue 4

⁴Lorenz et al., *In proceedings, Lecture Notes in Informatics, German Conference on Bioinformatics* 2009

Topics/Work of my PhD study

Published work

- ① ViennaRNA Package 2.0¹
- ② RNA/DNA hybrid structure prediction²
- ③ Secondary structure prediction with G-Quadruplexes³
- ④ **2D projections of secondary structure landscapes⁴**

Unpublished work

- ① **Detection of alternative low-free energy secondary structure states**
- ② **Prediction of (near) optimal, indirect RNA refolding paths**
- ③ **Coarse grained RNA folding kinetics with *ab-initio* partitioning**
- ④ **Simulation of cotranscriptional folding dynamics**

¹Lorenz et al., *Algorithms for Molecular Biology* 2011, 6:26

²Lorenz et al., *Bioinformatics* 2012 Vol. 28 no. 19

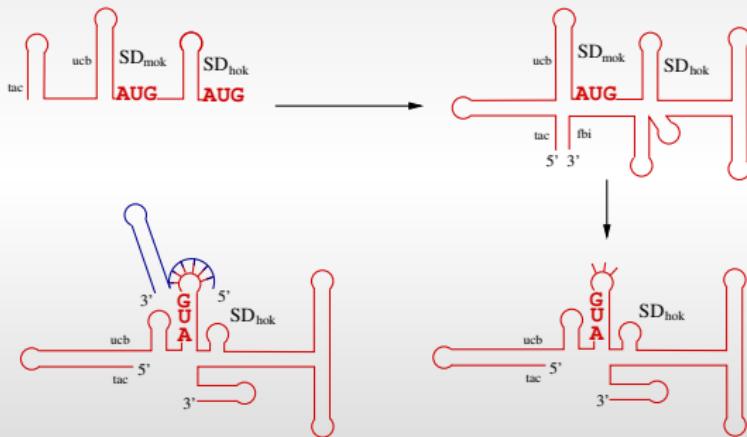
³Lorenz et al., *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 2013 Volume 10, Issue 4

⁴Lorenz et al., *In proceedings, Lecture Notes in Informatics, German Conference on Bioinformatics* 2009

Gene regulation by RNA structure changes

Cis-induced switches:

- transcriptional control (e.g. attenuators)
- translational control
 - co-transcriptional kinetic traps (e.g. sv11, ms2, hok/sok)



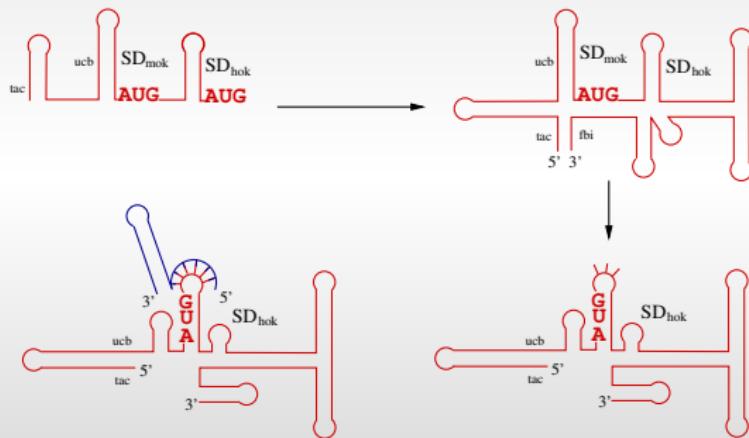
trans-induced switches:

- metabolite sensors (e.g. add, TPP, SAM)
- temperature sensors (e.g. bacterial virulence genes)

Gene regulation by RNA structure changes

Cis-induced switches:

- transcriptional control (e.g. attenuators)
- translational control
 - co-transcriptional kinetic traps (e.g. sv11, ms2, hok/sok)



trans-induced switches:

- metabolite sensors (e.g. add, TPP, SAM)
- temperature sensors (e.g. bacterial virulence genes)

RNA switch example⁵

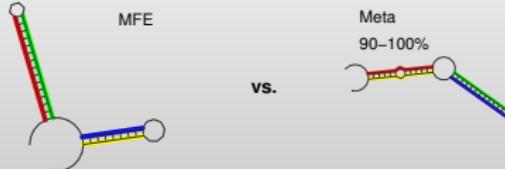
- folding path can be encoded within sequence
- utilization of helix length asymmetry
- two switches with almost identical energy landscape
- in vitro experiments to see co-transcriptional folding effect

5' – ABCD – 3' ('direct')



vs.

5' – DCBA – 3' ('indirect')



vs.

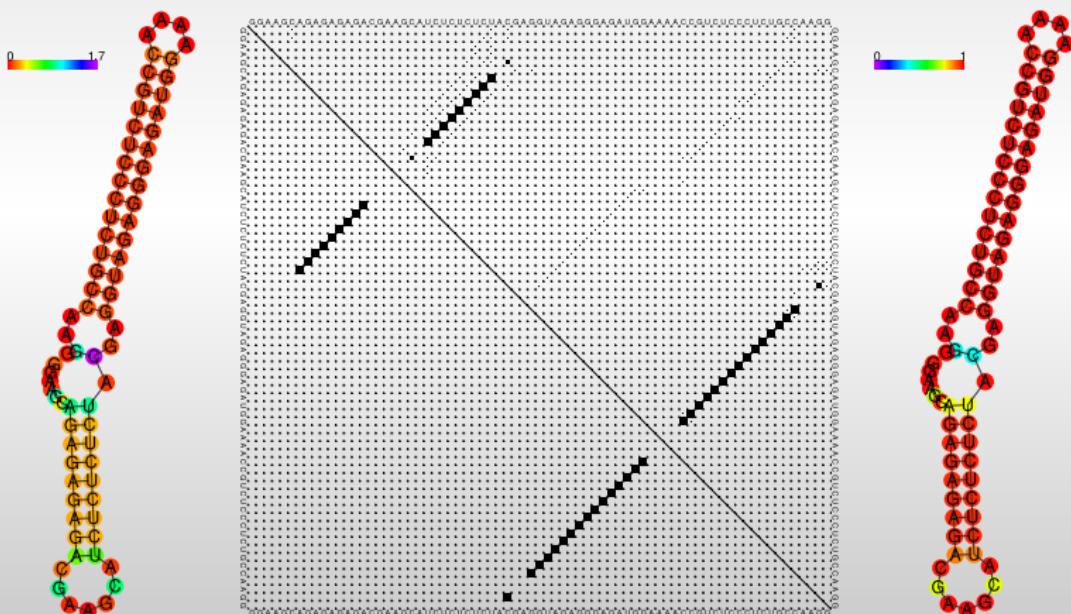
⁵Xayaphoummine et al. 2007

RNA switch example

- ① Thermodynamic equilibrium analysis
 - ▶ established tools
 - ▶ distance class partitioning
- ② Dynamic behavior of structural transitions
 - ▶ full length transcript
 - ▶ growing transcript

RNA switch example

```
$ RNAfold -p --MEA < xaya_dcba.2D
GGAAAGCAGAGAGACGAAGCAUCUCUCUACGAGGUAGAGGGAGAUGGAAAACCGUCUCCCUCUGCCAAGG
.....((((((.....))))))).((.....(((((((((.....))))))))))))...). (-41.90)
.....((((((.....))))))).|.((.....((((((.....))))))))))))...,[ -42.92]
.....((((((.....)))))).((.....((((((.....))))))))))))....{ -41.80 d=2.00}
.....((((((.....)))))).((.....((((((.....))))))))....{ -41.80 MEA=69.70}
frequency of mfe structure in ensemble 0.189913; ensemble diversity 3.34
```



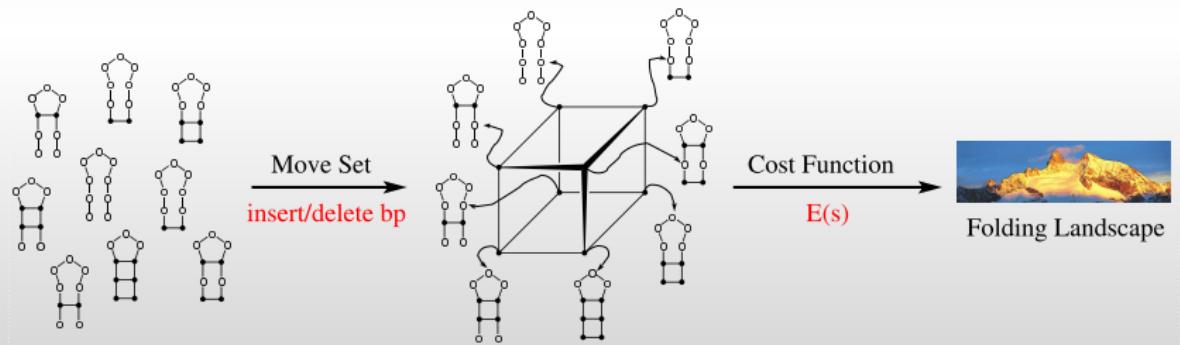
RNA switch example

```
$ RNAsubopt -s -e 3 < xaya_dcba.2D
GGAAGCAGAGAGACGAAGCAUCUCUCAUCAGGUAGAGGGAGAUGGAAAACCUCUCCUCUGCCAAGG -4190 300
.....(((((((((.....))))))),..(((((.....)))))))).... -41.90
.....(((((.....))))))..(((((.....)))))))).... -41.80
.....((((((.....)))))))....(((((.....)))))))).... -41.80
.....((((((.....)))))))..(((((.....)))))))).... -41.20
.....(((((.....))))))..(((((.....)))))))).... -41.20
.....(((((.....))))))..(((((.....)))))))).... -41.10
.....(((((.....)))))))....(((((.....)))))))).... -41.10
.....(((((.....)))))))....(((((.....)))))))).... -41.10
.....(((((.....)))))))..(((((.....)))))))).... -40.50
.....(((((.....))))))..(((((.....)))))))).... -40.50
.....(((((.....)))))))..(((((.....)))))))).... -40.40
.....(((((.....)))))))....(((((.....)))))))).... -40.40
.....(((((.....)))))))....(((((.....)))))))).... -40.40
.....(((((.....)))))))....(((((.....)))))))).... -40.00
.....(((((.....))))))..(((((.....)))))))).... -39.90
.....(((((.....))))))..(((((.....)))))))).... -39.80
.....(((((.....))))))..(((((.....)))))))).... -39.80
.....(((((.....))))))..(((((.....)))))))).... -39.70
((...))(((((.....))))))..(((((.....)))))))).... -39.60
((...))(((((.....))))))..(((((.....)))))))).... -39.50
....(((((.....))))))..(((((.....)))))))).... -39.20
.....(((((.....))))))..(((((.....)))))))).... -39.20
....(((((.....))))))..(((((.....)))))))).... -39.10
.....(((((.....))))))..(((((.....)))))))).... -39.10
.....(((((.....))))))..(((((.....)))))))).... -39.10
.....(((((.....))))))..(((((.....)))))))).... -39.10
....(((((.....))))))..(((((.....)))))))).... -39.00
....(((((.....))))))..(((((.....)))))))).... -39.00
.....(((((.....))))))..(((((.....)))))))).... -39.00
((...))(((((.....))))))..(((((.....)))))))).... -38.90
((...))(((((.....))))))..(((((.....)))))))).... -38.90
....(((((.....))))))..(((((.....)))))))).... -38.90
....(((((.....))))))..(((((.....)))))))).... -38.90
....(((((.....))))))..(((((.....)))))))).... -38.90
....(((((.....))))))..(((((.....)))))))).... -38.90
```

RNA secondary structure landscape

$$\text{Energy Landscape} = \{\Omega, \mathcal{M}, f\}$$

- Ω ... set of configurations (*RNA secondary structures*)
- \mathcal{M} ... move set/neighborhood relation on Ω (*insert/remove bp*)
- f ... fitness/energy function with $f : \Omega \rightarrow \mathbb{R}$ (*free energy*)

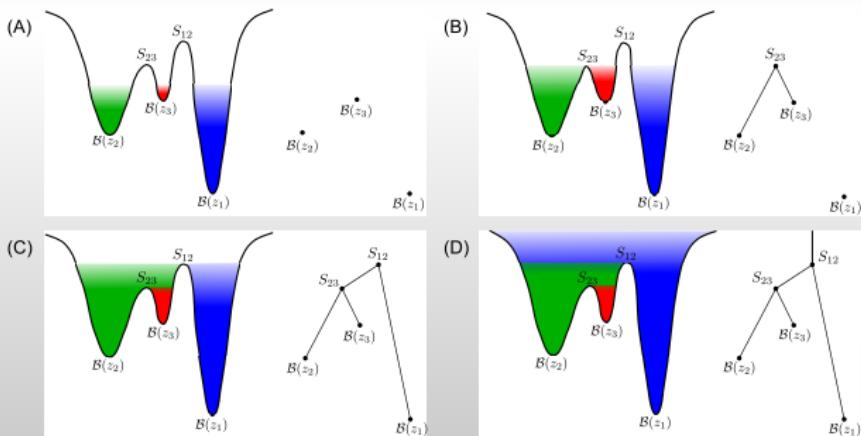


RNA secondary structure landscape

$$\text{Energy Landscape} = \{\Omega, \mathcal{M}, f\}$$

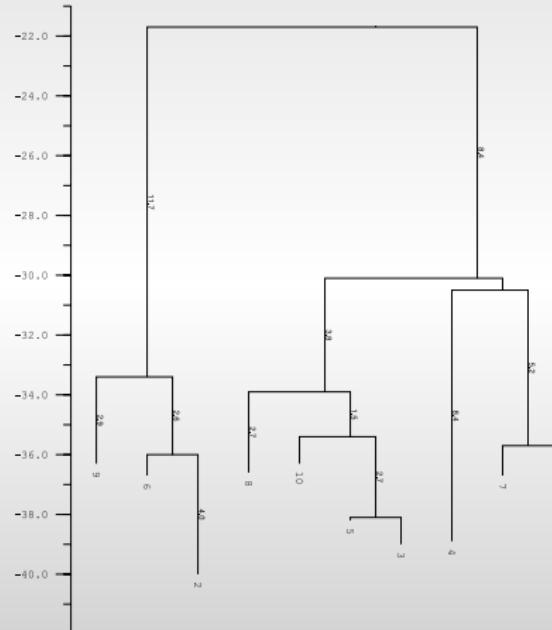
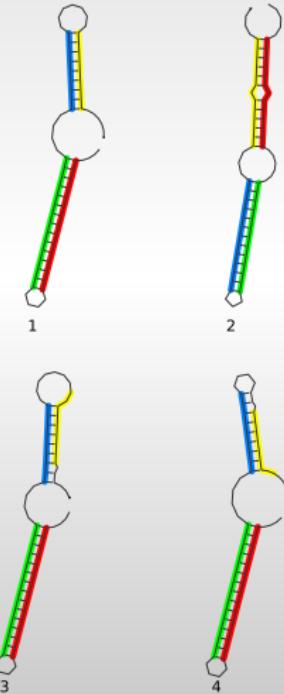
- Ω ... set of configurations (*RNA secondary structures*)
- \mathcal{M} ... move set/neighborhood relation on Ω (*insert/remove bp*)
- f ... fitness/energy function with $f : \Omega \rightarrow \mathbb{R}$ (*free energy*)

Coarse graining with Barrier trees



RNA switch example

- ① Exhaustively enumerate structures (RNAsubopt^6)
- ② Produce landscape representation (barriers⁷)



⁶Wuchty et al., 1999

⁷Wolfinger et al., 2004

Distance class partitioning

- nice way to project high-dimensional energy landscape into lower dimensions
- classified dynamic programming approach
- select n reference structures s_1, s_2, \dots (of interest)
- lump structures into partitions according their base pair distance to the references
- compute the MFE, partition function for each distance class

Implementation examples:

- ① RNAbor⁸ (one reference structure)
- ② RNA2Dfold⁹ (two reference structures)

⁸Freyhult et al, 2007

⁹Lorenz et al., 2009

$$\begin{aligned}
 F_{i,j}^{\kappa,\lambda} &= \min \left\{ \begin{array}{l} F_{i,j-1}^{\kappa-\delta_1^1(i,j), \lambda-\delta_1^2(i,j)}, \\ \min_{\substack{i \leq u < j \\ \omega_1 + \hat{\omega}_1 = \kappa - \delta_1^1(i,j,u) \\ \omega_2 + \hat{\omega}_2 = \lambda - \delta_2^2(i,j,u)}} F_{i,u-1}^{\omega_1, \omega_2} + C_{u,j}^{\hat{\omega}_1, \hat{\omega}_2} \end{array} \right. \\
 C_{i,j}^{\kappa,\lambda} &= \min \left\{ \begin{array}{l} \mathfrak{H}(i, j, \kappa, \lambda), \\ \min_{i < p < q < j} \left\{ C_{p,q}^{\kappa-\delta_3^1(i,j,p,q), \lambda-\delta_3^2(i,j,p,q)} + \mathcal{I}(i, j, p, q) \right\}, \\ \min_{\substack{i < u < j \\ \omega_1 + \hat{\omega}_1 = \kappa - \delta_4^1(i,j,u) \\ \omega_2 + \hat{\omega}_2 = \lambda - \delta_4^2(i,j,u)}} \left\{ M_{i+1,u}^{\omega_1, \omega_2} + \hat{M}_{u+1,j-1}^{\hat{\omega}_1, \hat{\omega}_2} + a \right\} \end{array} \right. \\
 M_{i,j}^{\kappa,\lambda} &= \min \left\{ \begin{array}{l} M_{i,j}^{\kappa-\delta_1^1(i,j-1), \lambda-\delta_1^2(i,j)} + c \\ \min_{i \leq u < j} \left\{ (u-i) \cdot c + C_{u,j}^{\kappa-\delta_5^1(i,j,u), \lambda-\delta_5^2(i,j,u)} + b \right\}, \\ \min_{\substack{i \leq u < j \\ \omega_1 + \hat{\omega}_1 = \kappa - \delta_2^1(i,j,u) \\ \omega_2 + \hat{\omega}_2 = \lambda - \delta_2^2(i,j,u)}} \left\{ M_{i,u-1}^{\omega_1, \omega_2} + C_{u,j}^{\hat{\omega}_1, \hat{\omega}_2} + b \right\}, \end{array} \right. \\
 \hat{M}_{i,j}^{\kappa,\lambda} &= \min \left\{ \begin{array}{l} C_{i,j}^{\kappa,\lambda} + b \\ \hat{M}_{i,j-1}^{\kappa-\delta_1^1(i,j), \lambda-\delta_1^2(i,j)} + c, \end{array} \right.
 \end{aligned}$$

with:

$$\begin{aligned}
 \delta_1^x(i, j) &= d_{\text{BP}}(s_x[i, j], s_x[i, j-1]) \\
 \delta_2^x(i, j, u) &= d_{\text{BP}}(s_x[i, j], s_x[i, u-1] \cup s_x[u, j]) \\
 \delta_3^x(i, j, p, q) &= d_{\text{BP}}(s_x[i, j], \{(i, j)\} \cup s_x[p, q]) \\
 \delta_4^x(i, j, u) &= d_{\text{BP}}(s_x[i, j], \{(i, j)\} \cup s_x[i+1, u] \cup s_x[u+1, j-1]) \\
 \delta_5^x(i, j, u) &= d_{\text{BP}}(s_x[i, j], s_x[u, j])
 \end{aligned}$$

RNA2Dfold in a nutshell

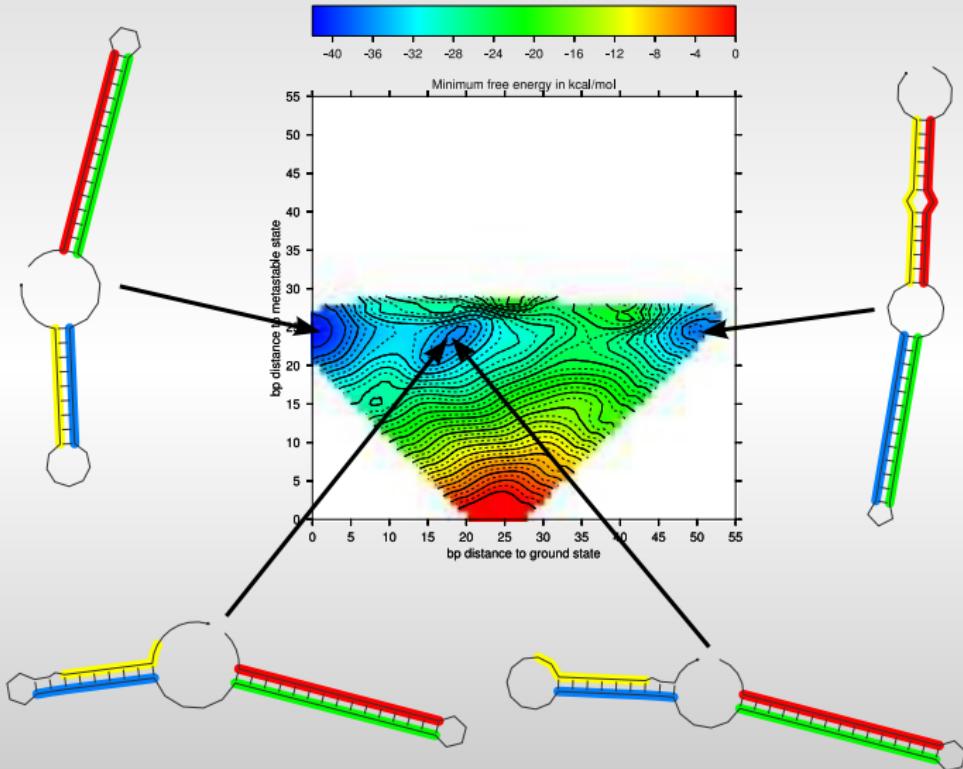
$$\text{MFE}_{\kappa, \lambda} = \min_{\substack{s \in S \\ \delta(s, s_1) = \kappa \\ \delta(s, s_2) = \lambda}} E(s)$$

Algorithmic details

- MFE, Partition function, Stochastic backtracking
- Asymptotic time complexity $O(n^7)$, memory $O(n^4)$
- Speedup through
 - ▶ exploitation of sparseness
 - ▶ parallel computation
 - ▶ κ, λ may be limited to maximum distance
- Results within reasonable time for sequences of up to 400 nt
- Part of the ViennaRNA Package since version 2

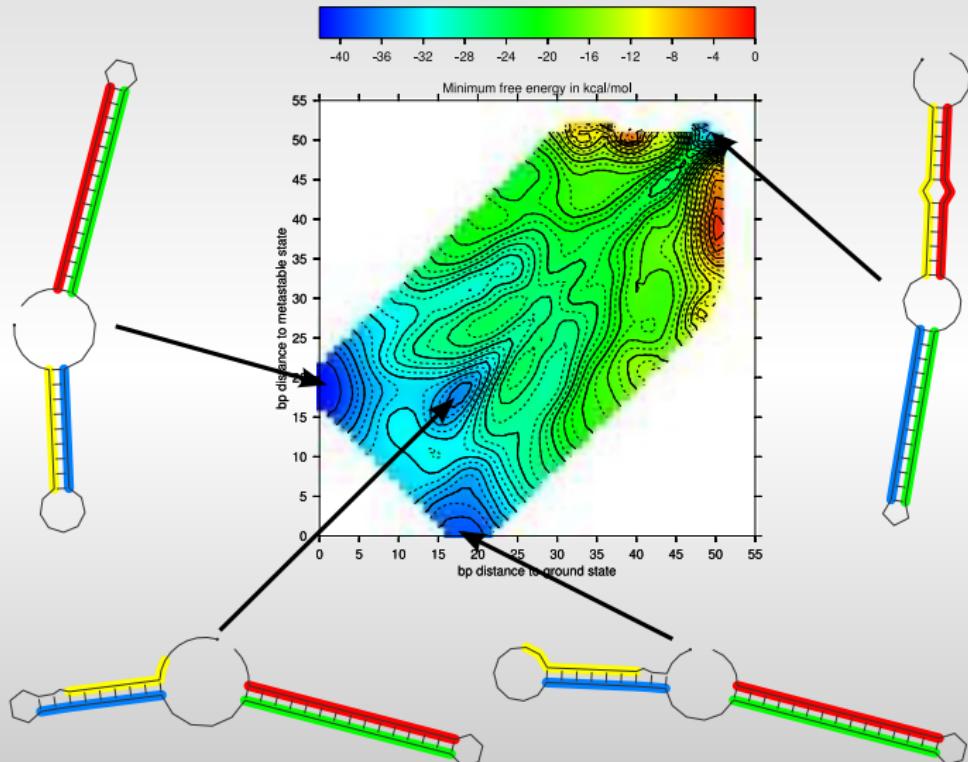
RNA2Dfold - search for meta stable states

MFE structure and unfolded state as references



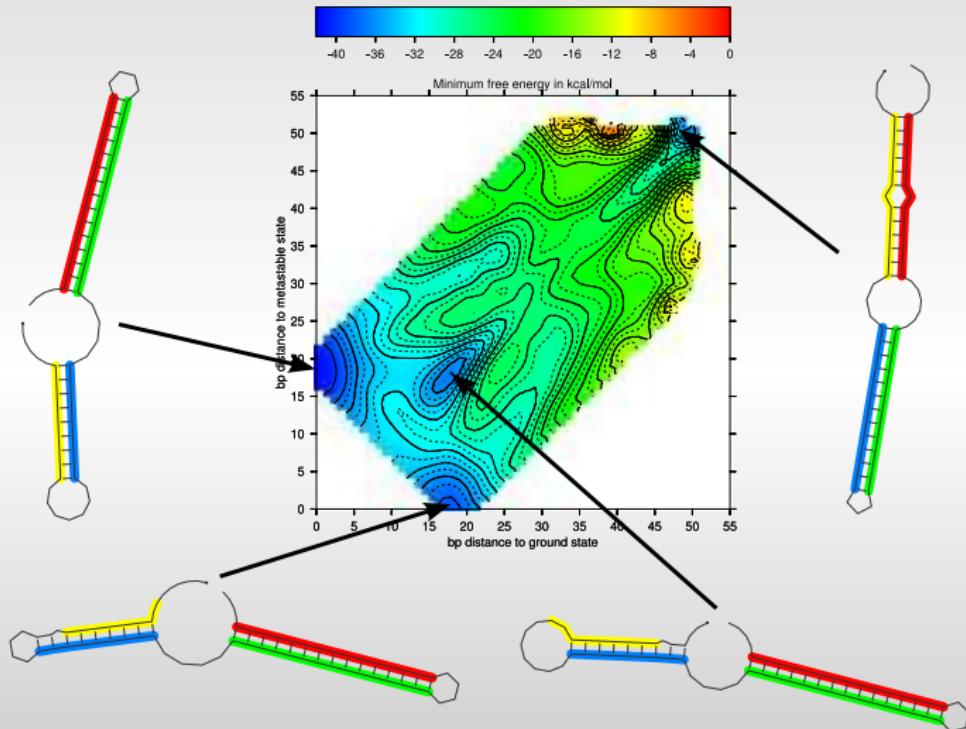
RNA2Dfold - search for meta stable states

MFE structure and first shifted structure as references



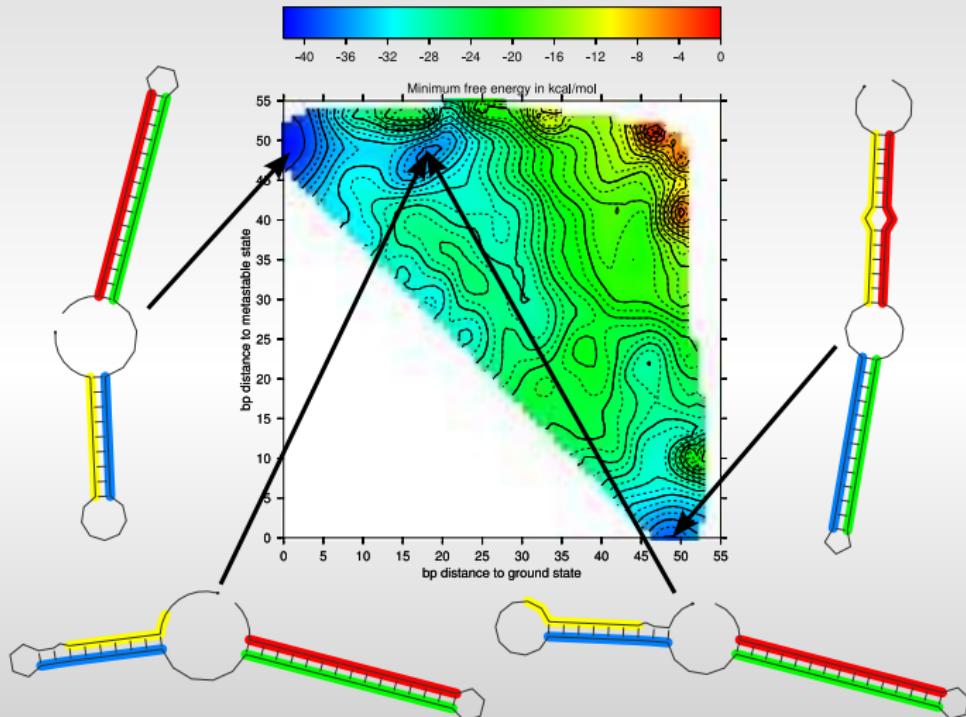
RNA2Dfold - search for meta stable states

MFE structure and second shifted structure as references



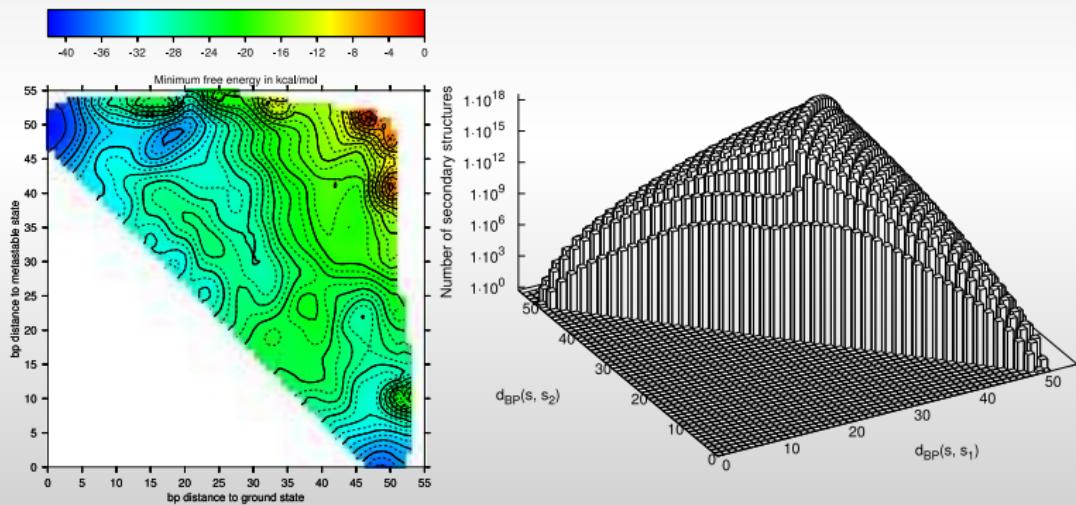
RNA2Dfold - search for meta stable states

MFE structure and meta stable structure as references



RNA2Dfold - distance class population

The secondary structure space is **huge!**



RNA structure refolding

How difficult is refolding from one structure into another?



RNA structure refolding

How difficult is refolding from one structure into another?



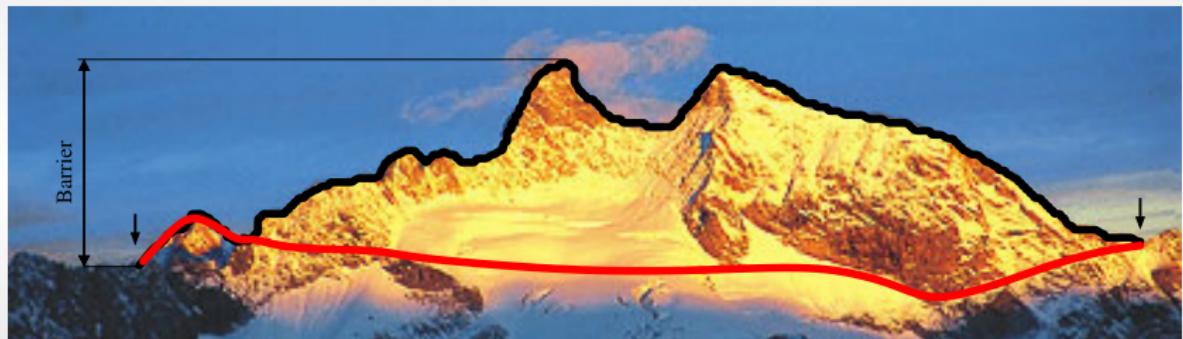
RNA structure refolding

How difficult is refolding from one structure into another?



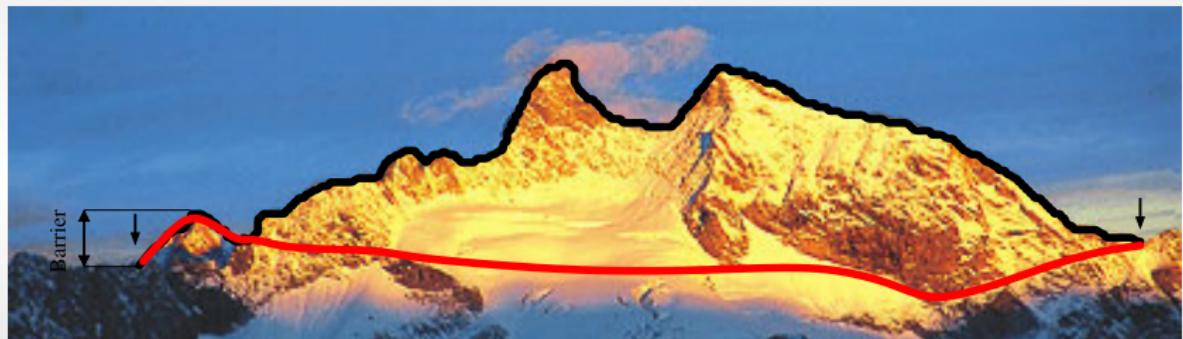
RNA structure refolding

How difficult is refolding from one structure into another?



RNA structure refolding

How difficult is refolding from one structure into another?



RNA structure refolding

How difficult is refolding from one structure into another?



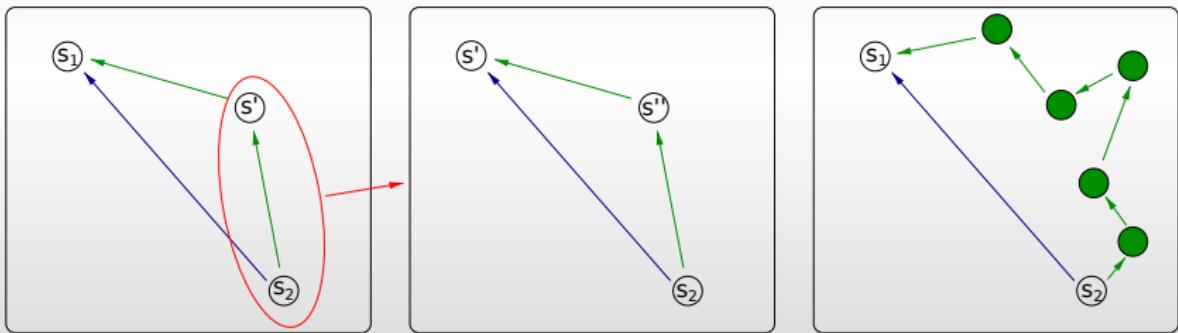
Finding a refolding path with lowest barrier is *NP-hard* even for direct paths!¹⁰

Heuristics: Morgan-Higgs, findpath, RNAtabupath, ...

¹⁰Manuch et al. 2009

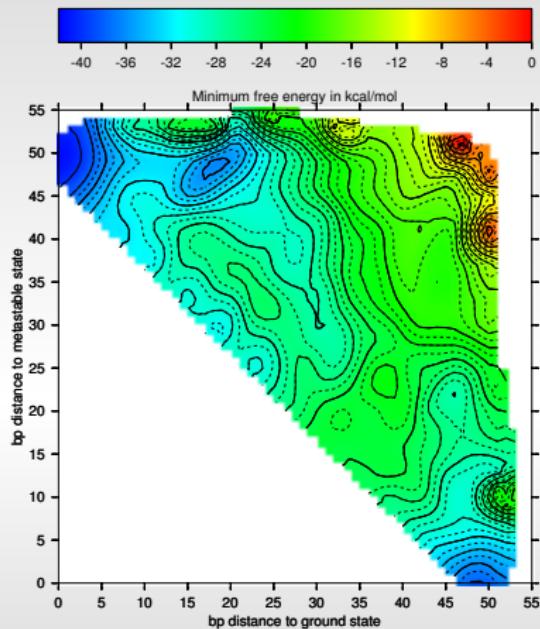
Construction of indirect refolding paths with Pathfinder¹¹

- compute direct refolding path between s_1 and s_2
- generate stabilizing point s'
- compute direct refolding path between $s_{1,2}$ and s'
- accept new path if it yields lower energy barrier
- iteratively refine resulting path if necessary

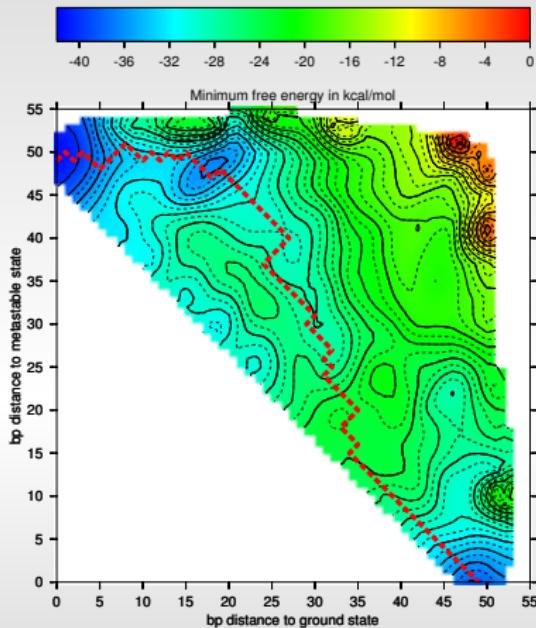


¹¹Lorenz et al., 2009

Lowest energy barrier with indirect refolding paths

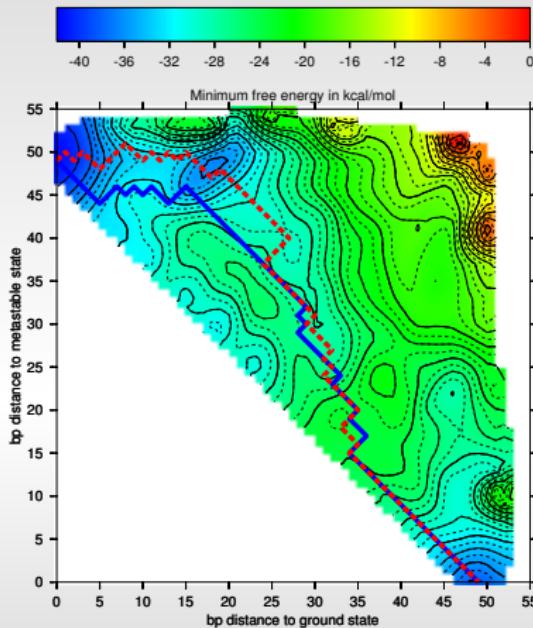


Lowest energy barrier with indirect refolding paths



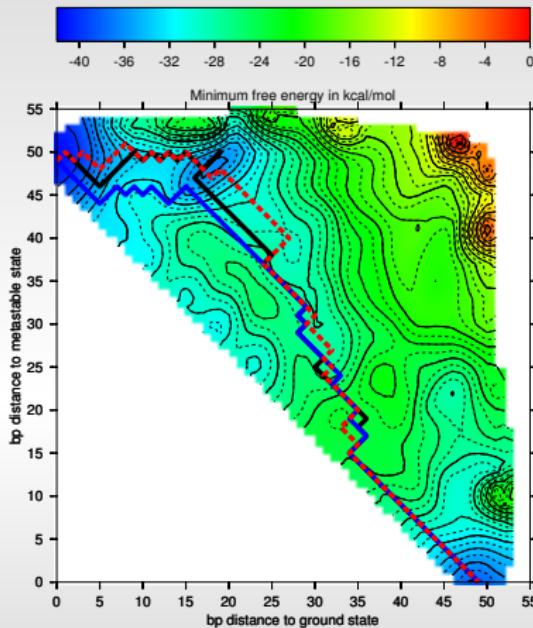
- Estimate using RNA2Dfold

Lowest energy barrier with indirect refolding paths



- Estimate using RNA2Dfold
- Pathfinder approach using stabilizing points from RNA2Dfold

Lowest energy barrier with indirect refolding paths



- Estimate using RNA2Dfold
- Pathfinder approach using stabilizing points from RNA2Dfold
- Exhaustive search

How to take the dynamical behavior of refolding into account?

- ① Stochastic single trajectory methods (Monte Carlo)
- ② Deterministic whole system approaches (Markov process)

How to take the dynamical behavior of refolding into account?

- ① Stochastic single trajectory methods (Monte Carlo)
- ② **Deterministic whole system approaches (Markov process)**
 - ▶ General approach
 - ▶ Reduction of the state space
 - ▶ Sampling of the state space

RNA folding kinetics as a Markov process

Population density of structure states over time

$$\frac{dP_i(t)}{dt} = \sum_{j \neq i} (P_j(t)k_{ji} - P_i(t)k_{ij})$$

Equilibrium condition $\frac{dP_i(t)}{dt} = 0 \rightarrow P_j(t)k_{ji} = P_i(t)k_{ij}$

$$\frac{k_{ji}}{k_{ij}} = \frac{e^{-\beta E(s_i)}}{e^{-\beta E(s_j)}} = e^{\beta(E(s_j) - E(s_i))} = e^{\beta \Delta G_{ij}}$$

Transition rates between the structures, e.g. Metropolis rule

$$k_{ij} = \begin{cases} \tau_0^{-1} e^{-\beta \Delta G_{ij}} & \text{if } E(s_j) > E(s_i), \\ \tau_0^{-1} & \text{otherwise} \end{cases}$$

Master equation with $R = \{k_{ij}\}$

$$\frac{d}{dt} \vec{p}(t) = R \vec{p}(t) \quad \rightarrow \quad \vec{p}(t) = \vec{p}(0) \cdot e^{t \cdot R}$$

RNA folding kinetics as a Markov process with macro states

Partitioning of the state space

$$\mathcal{A} = \{\alpha_1, \alpha_2, \dots\} \text{ with } \Omega = \cup_j \alpha_j$$

Major assumption for macro state kinetics

- any macro state is in equilibrium itself, i.e.

$$Prob[s_i|\alpha] = \frac{e^{-\beta E(s_i)}}{Q_\alpha}$$

Therefore, the transition rate is

$$k_{\alpha_i, \alpha_j} = \sum_{s_x \in \alpha_i} \sum_{\substack{s_y \in \alpha_j \\ \alpha_i \neq \alpha_j}} k_{xy} Prob[s_x|\alpha_i]$$

$$k_{\alpha_i, \alpha_j} = \frac{1}{Q_{\alpha_i}} \sum_{s_x \in \alpha_i} \sum_{s_y \in \alpha_j} k_{xy} e^{-\beta E(s_x)}$$

RNA folding kinetics as a Markov process with macro states

Partitioning of the state space

$$\mathcal{A} = \{\alpha_1, \alpha_2, \dots\} \text{ with } \Omega = \cup_j \alpha_j$$

Major assumption for macro state kinetics

- any macro state is in equilibrium itself, i.e.

$$Prob[s_i|\alpha] = \frac{e^{-\beta E(s_i)}}{Q_\alpha}$$

Therefore, the transition rate is

$$k_{\alpha_i, \alpha_j} = \sum_{s_x \in \alpha_i} \sum_{\substack{s_y \in \alpha_j \\ \alpha_i \neq \alpha_j}} k_{xy} Prob[s_x|\alpha_i]$$

$$k_{\alpha_i, \alpha_j} = \frac{1}{Q_{\alpha_i}} \sum_{s_x \in \alpha_i} \sum_{s_y \in \alpha_j} k_{xy} e^{-\beta E(s_x)}$$

Exhaustive approach!

RNA folding kinetics as a Markov process with macro states

Sampling approach

$$\text{Requirement: } p(s_x \in S_\alpha) = \frac{\exp(-\beta E(s_x))}{Q_\alpha}$$

Therefore, the transition rate can be approximated

$$\hat{k}_{\alpha_1, \alpha_2} \approx \frac{1}{|S_{\alpha_1}|} \sum_{s_i \in S_{\alpha_1}} \sum_{\substack{s_j \in \alpha_2 \\ s_j \in \mathcal{N}(s_i)}} k_{ij}$$

$$\tilde{k}_{\alpha_1, \alpha_2} \approx \frac{1}{|S_{\alpha_2}|} \sum_{s_j \in S_{\alpha_2}} \sum_{\substack{s_i \in \alpha_1 \\ s_i \in \mathcal{N}(s_j)}} \frac{\pi_{\alpha_2}}{\pi_{\alpha_1}} k_{ji}$$

$$k_{\alpha_1, \alpha_2} \approx \frac{|S_{\alpha_1}| \cdot \hat{k}_{\alpha_1, \alpha_2} + |S_{\alpha_2}| \cdot \tilde{k}_{\alpha_1, \alpha_2}}{|S_{\alpha_1}| + |S_{\alpha_2}|}$$

$$\approx \frac{1}{|S_{\alpha_1}| + |S_{\alpha_2}|} \left\{ \sum_{s_i \in S_{\alpha_1}} \sum_{\substack{s_j \in \alpha_2 \\ s_j \in \mathcal{N}(s_i)}} k_{ij} + \sum_{s_j \in S_{\alpha_2}} \sum_{\substack{s_i \in \alpha_1 \\ s_i \in \mathcal{N}(s_j)}} \frac{\pi_{\alpha_2}}{\pi_{\alpha_1}} k_{ji} \right\}$$

RNA folding kinetics as a Markov process with macro states

Sampling approach

$$\text{Requirement: } p(s_x \in S_\alpha) = \frac{\exp(-\beta E(s_x))}{Q_\alpha}$$

Therefore, the transition rate can be approximated

$$\hat{k}_{\alpha_1, \alpha_2} \approx \frac{1}{|S_{\alpha_1}|} \sum_{s_i \in S_{\alpha_1}} \sum_{\substack{s_j \in \alpha_2 \\ s_j \in \mathcal{N}(s_i)}} k_{ij}$$

$$\tilde{k}_{\alpha_1, \alpha_2} \approx \frac{1}{|S_{\alpha_2}|} \sum_{s_j \in S_{\alpha_2}} \sum_{\substack{s_i \in \alpha_1 \\ s_i \in \mathcal{N}(s_j)}} \frac{\pi_{\alpha_2}}{\pi_{\alpha_1}} k_{ji}$$

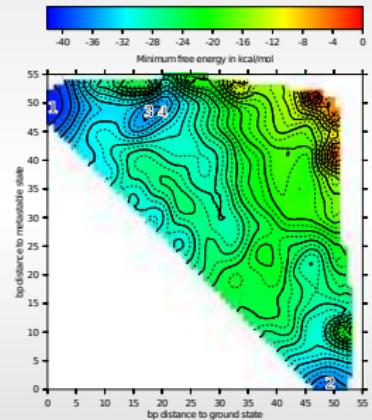
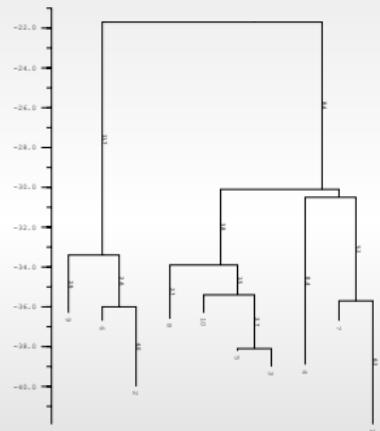
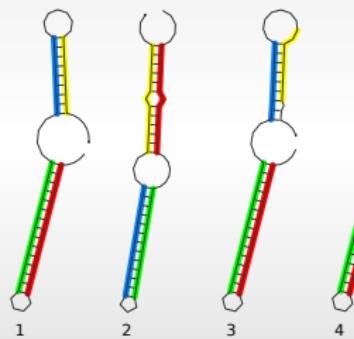
$$k_{\alpha_1, \alpha_2} \approx \frac{|S_{\alpha_1}| \cdot \hat{k}_{\alpha_1, \alpha_2} + |S_{\alpha_2}| \cdot \tilde{k}_{\alpha_1, \alpha_2}}{|S_{\alpha_1}| + |S_{\alpha_2}|}$$

$$\approx \frac{1}{|S_{\alpha_1}| + |S_{\alpha_2}|} \left\{ \sum_{s_i \in S_{\alpha_1}} \sum_{\substack{s_j \in \alpha_2 \\ s_j \in \mathcal{N}(s_i)}} k_{ij} + \sum_{s_j \in S_{\alpha_2}} \sum_{\substack{s_i \in \alpha_1 \\ s_i \in \mathcal{N}(s_j)}} \frac{\pi_{\alpha_2}}{\pi_{\alpha_1}} k_{ji} \right\}$$

Use distance class partitioning for \mathcal{A}

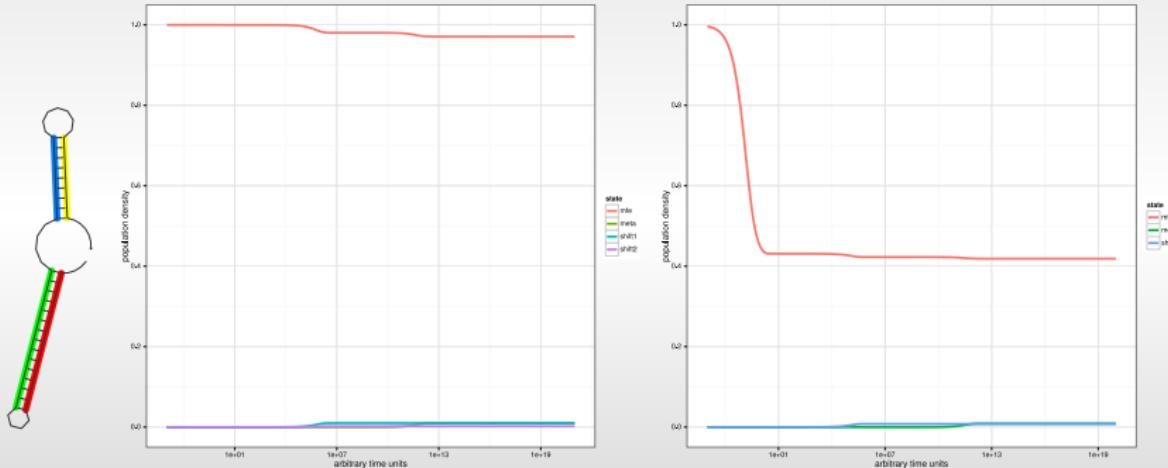
RNA folding kinetics (switch example)

Comparison between Gradient basin approach and 2DKin



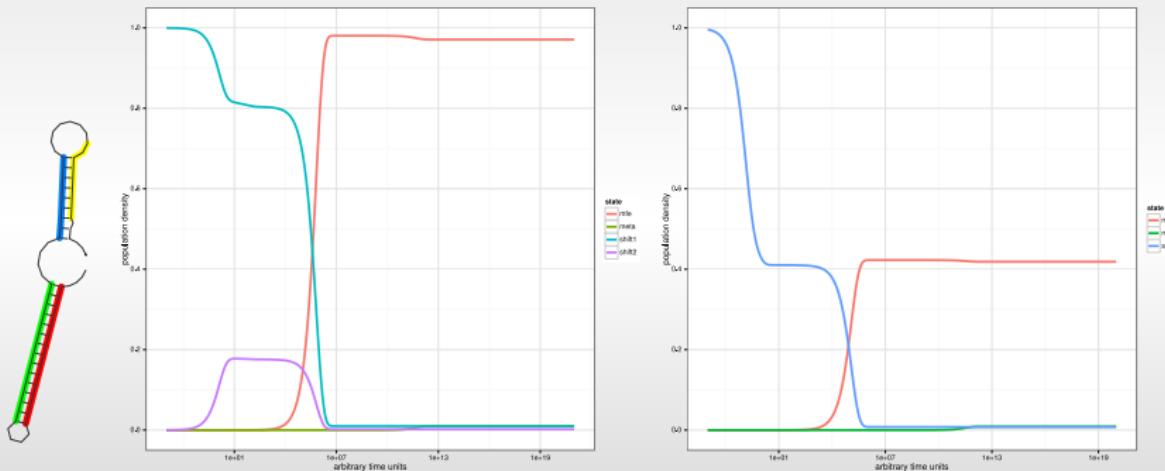
RNA folding kinetics (switch example)

starting with 100% MFE structure



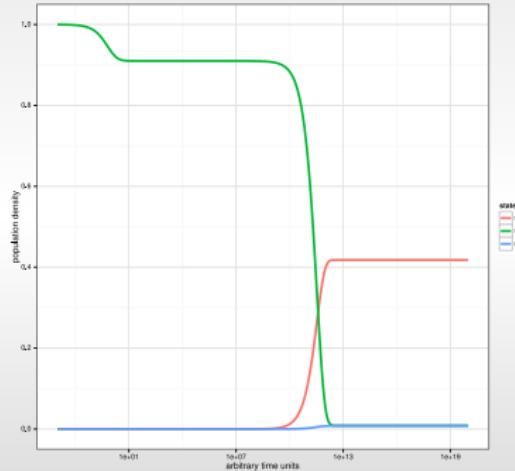
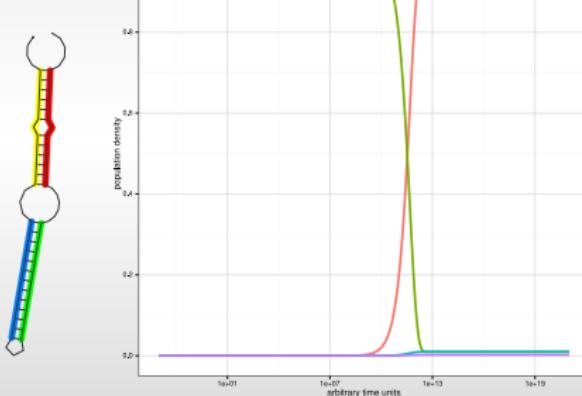
RNA folding kinetics (switch example)

starting with 100% shifted structure



RNA folding kinetics (switch example)

starting with 100% meta stable structure

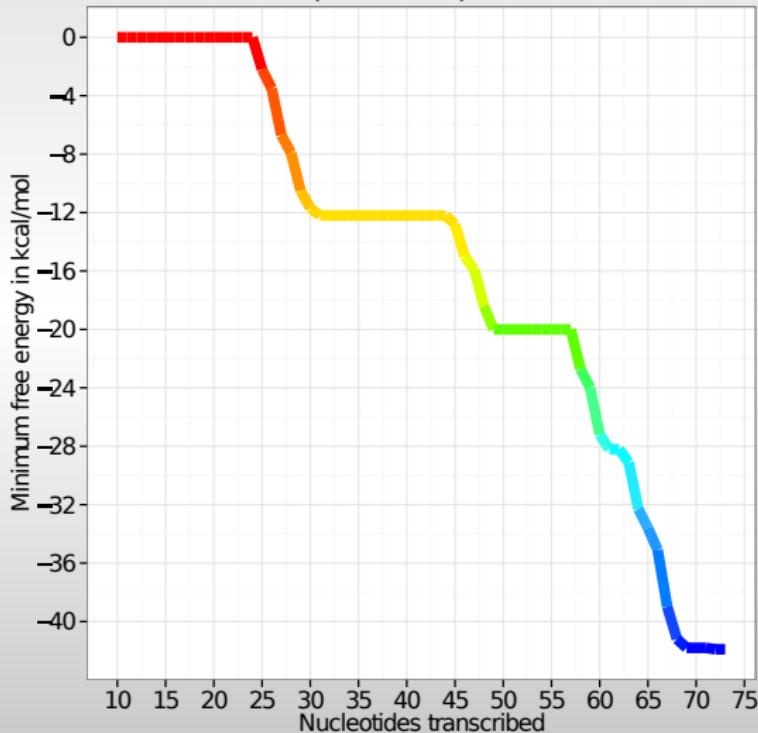


Folding dynamics on varying landscapes

- ① Thermodynamic equilibrium analysis
- ② Mapping of consecutive folding kinetics simulations

Co-transcriptional folding (equilibrium analysis)

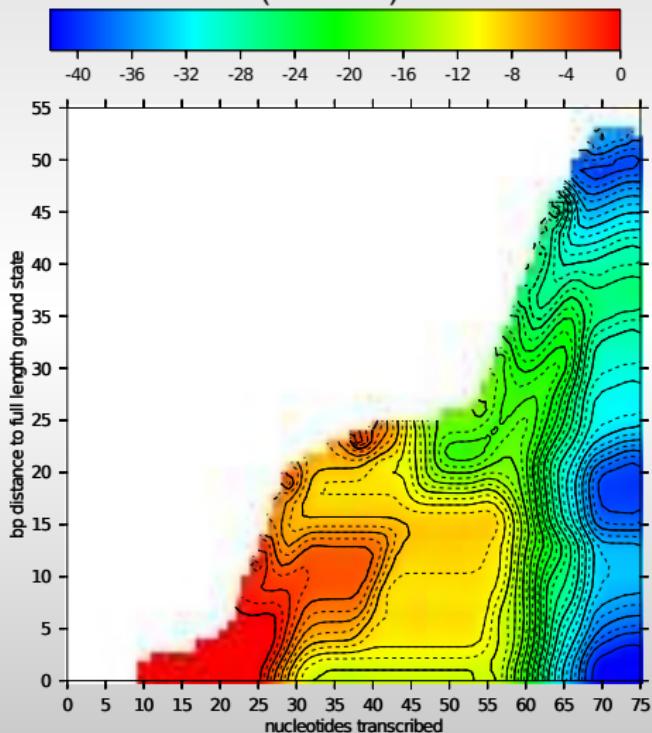
Computing MFEs for each transcription step
(RNAfold)



Co-transcriptional folding (equilibrium analysis)

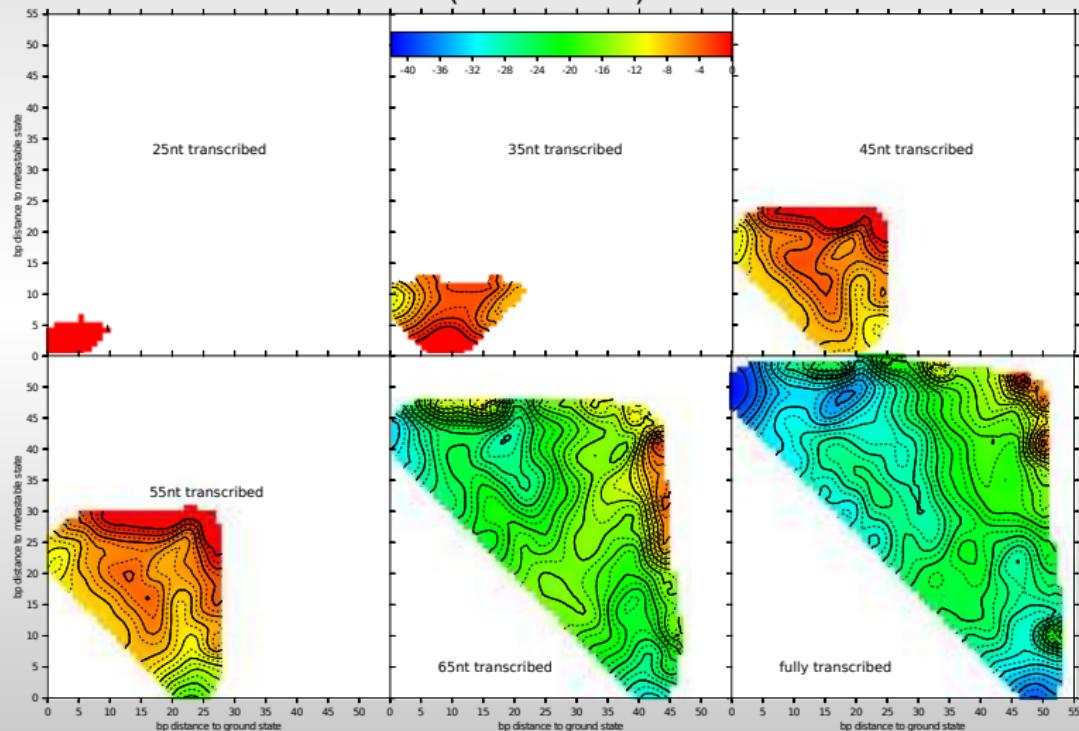
Computing MFEs for 1D distance classes for each transcription step

(RNAbor)



Co-transcriptional folding (equilibrium analysis)

Computing MFEs for 2D distance classes for each transcription step
(RNA2Dfold)



Co-transcriptional folding (equilibrium analysis)

Computing MFEs for 2D distance classes for each transcription step
(RNA2Dfold)

MOVIE!

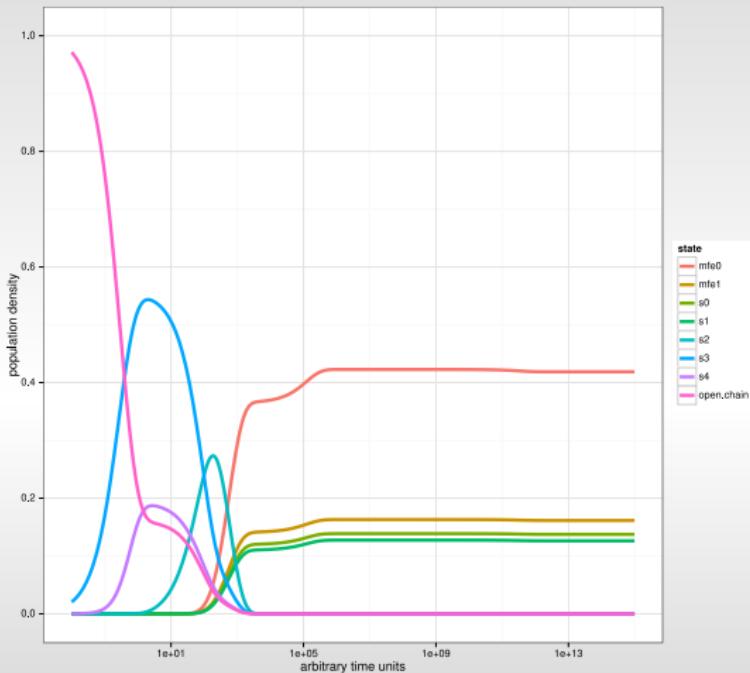
Co-transcriptional folding (refolding dynamics)

barmap¹² like approach:

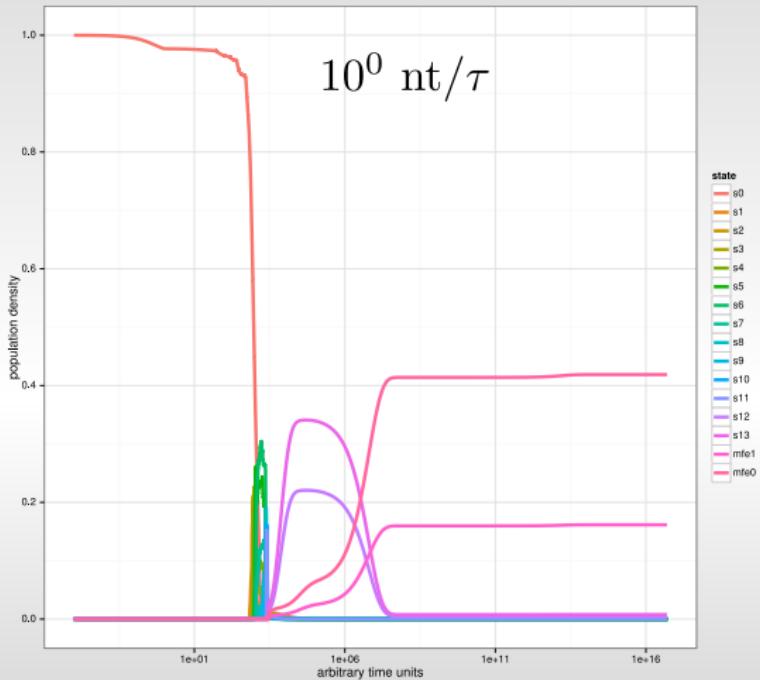
- compute rates for each landscape
- select initial population density
- map the states from one landscape onto another:
 - ▶ variable RNA chain length:
 $\text{seq}[1 : i] \rightarrow \text{seq}[1 : i + 1]$:
 $c_{\kappa, \lambda} \rightarrow c_{\kappa+\alpha, \lambda+\beta}$ with
 $\alpha = (\text{ref1}[1 : i + 1] - \text{ref1}[1 : i])$ and
 $\beta = (\text{ref2}[1 : i + 1] - \text{ref2}[1 : i])$
- consecutive simulations initialized with population density of previous one

¹²Hofacker et al. 2010

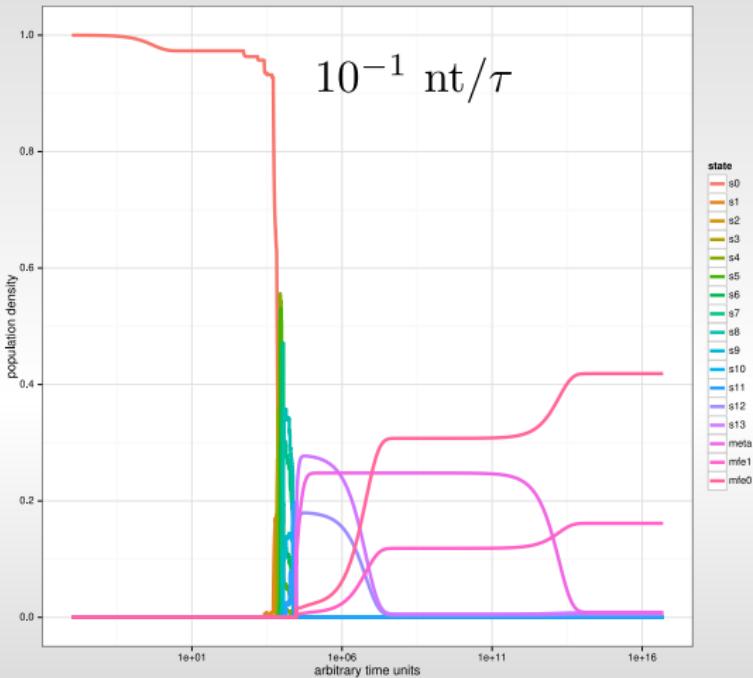
Co-transcriptional folding (refolding dynamics)



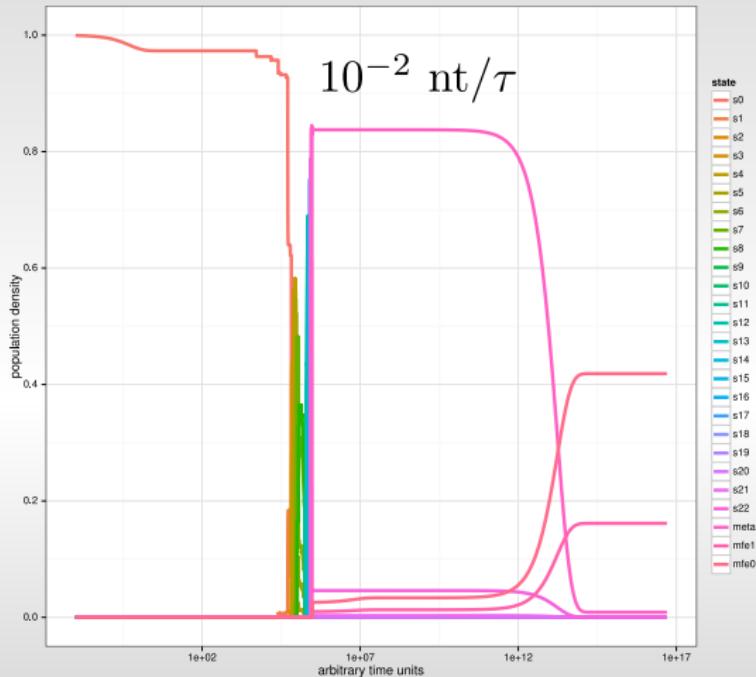
Co-transcriptional folding (refolding dynamics)



Co-transcriptional folding (refolding dynamics)



Co-transcriptional folding (refolding dynamics)



Conclusions and Outlook

- Sampling from distance classes can be used for kinetic simulation
- Longer molecules can be analyzed
- Kinetics on varying landscapes

Future perspectives

- Relationship of transcription times and arbitrary time units
- Regulation of transcription speed through RNA/DNA hybrids
- Replacement of (slow) classified DP approach by faster sampling approaches
- Extension to *trans*-acting switches
- Incorporation of higher-order structures

Thanks to

- Andrea Tanzer ❤
- Ivo L. Hofacker
- Christoph Flamm
- Michael T. Wolfinger
- Christian Höner zu Siederdissen
- Entire TBI team
- My family and friends

Thank you for your attention!