Sampling strategies to approximate RNA folding kinetics

Gregor Entzian

University of Vienna, Faculty of Chemistry, Department of Theoretical Chemistry

entzian@tbi.univie.ac.at

Supervisor: Univ.-Prof. Dipl.-Phys. Dr. Ivo L. Hofacker

February 14, 2017
RNA Switch Folding Kinetics

[Graph showing population density over time with RNA sequences highlighting ground state and alternative state]

Ground state  "On"

Alternative state  "Off"
\[ L = \{S, f, M\}; \quad S = \{s_1, s_2, \ldots\}; \quad f : s \rightarrow \mathbb{R}; \quad M = " + / − 1bp" \]
growth: $1.8^N$

atoms in the universe: $10^{80}$

$1.8^N = 10^{80} \rightarrow N = 312$ nucleotides
Coarse Graining

Energy

Conformation space

macro state 2/
basin 2

C_{1,2}

macro state 1/
basin 1

C_{1,3}

macro state 3/
basin 3

C_{3,4}

macro state 4/
basin 4

contact surface

barrier 2,3

gradient walk

saddle point

C_{1,2} C_{1,3} C_{3,4}

Energy

Conformation space
RNA Folding Kinetics

\[ \frac{dp_i(t)}{dt} = \sum_{j \neq i} [p_j(t) r_{ji} - p_i(t) r_{ij}] \]
Goals

- Make RNA folding kinetics prediction applicable to biologically relevant sequence lengths (> 100nt)

How to do that?

1. Sample the energy landscape to cover most important states
2. Determine physically meaningful partitioning of partial landscape into macro states
3. Derive good transition rate approximations for resulting macro states
4. Implement the above into a pipeline
Exhaustive enumeration
- up to a certain threshold
- exponential number of structures
→ only for short sequences

Boltzmann sampling
(importance sampling)

\[ P(s) = \frac{e^{-\frac{E(s)}{RT}}}{Q} \] with
\[ Q = \sum_s e^{-\frac{E(s)}{RT}} \]
- highly redundant
- only structures with small energy deviations from the ground state

Variable temperature sampling

\[ T = \xi \cdot T_0, \xi > 1 \]
- undirected
- large \( T = \text{sampling from uniform distribution} \)
New Approach: Guiding Potentials

- sampling with focus on important reference structures

\[ p(s_1) = p(s_2) = p(s_{mfe}) \]

\[ E'(s) = E(s) + \hat{E}(s) \]

\[ \hat{E}(s) = d(s, s_1) \cdot w_1 + d(s, s_2) \cdot w_2 \]
Interactive Guided Sampling - Iteration 1

http://www.tbi.univie.ac.at/~entzian/2DPlotter

Selected structures:

Y-reference:

X-reference:
Interactive Guided Sampling - New References

http://www.tbi.univie.ac.at/~entzian/2DPlotter
Interactive Guided Sampling - Iteration 2

http://www.tbi.univie.ac.at/~entzian/2DPlotter

[Image of RNA structures and graphs]

Gregor Entzian (TBI)
Tasks

1. Develop an automated iterative **sampling strategy** which is fast and produces both, diverse and most important structures of the energy landscape.

2. Develop **cluster-strategies** for RNA structures, to identify important structures and to generate macrostates.

3. Develop methods for computing the **transition rates** for incomplete landscapes.

4. Construct a **pipeline and programs** to compute RNA folding kinetics for long sequences and explore the underlying energy landscape. Implement a **web server**, which provides a comfortable graphical user interface.
Thank you!

- Ivo Hofacker
- RNALands project team:
  - Andrea Tanzer, Ronny Lorenz, Maria Waldl, Yann Ponty, Mireille Regnier, Hélène Touzet, Loïc Paulevé, Alain Denise, Juraj Michalik