Efficient approximations of RNA kinetics landscape using non-redundant sampling

Juraj Michalik¹ Helene Touzet² Yann Ponty¹

¹ - Equipe AMIB, Laboratoire LIX - Inria Saclay, France

² - Equipe Bonsai, Laboratoire CRIStAL - Université Lille 1, France

July 24, 2018





Image : Lorenz et al, GCB'09

J. Michalik, H. Touzet, Y. Ponty

Non-redundant sampling



Landscape : Lorenz et al, GCB'09, Structures : Varna - Darty et al, Bioinf. (2009)















Sampling of local minima



• Ragged nature of RNA landscapes \rightarrow Sampling representative set of LMs

Sampling of local minima



Ragged nature of RNA landscapes → Sampling representative set of LMs
Sampling according to Boltzmann distribution:

$$Prob(S) = rac{e^{rac{-\Delta G}{RT}}}{\mathcal{Z}}$$

- R: Boltzmann constant
- T: Absolute temperature
- \mathcal{Z} : Partition function

Sampling of local minima



Ragged nature of RNA landscapes → Sampling representative set of LMs
Sampling according to Boltzmann distribution:

$$Prob(S) = rac{e^{rac{-\Delta G}{RT}}}{\mathcal{Z}}$$

- *R*: Boltzmann constant
- T: Absolute temperature
- \mathcal{Z} : Partition function
- Non-redundant sampling \rightarrow Access those LMs having higher free-energy (ΔG)

Effect of Redundancy



Effect of Redundancy



J. Michalik, H. Touzet, Y. Ponty



Local optimality ensured by saturation



Local optimality ensured by saturation



Local optimality ensured by saturation



 ● Flat structures = Alternative local motifs within local minima
→ Modular decomposition (DP)

 Combinatorial LMs, Boltzmann-distributed wrt Turner energy model (cf. Waldispühl *et al*, J. Comput. Biol. 2007)





A, B - Subspaces of LMs featuring given motifs $\subset S$



A can be separated to subspaces

J. Michalik, H. Touzet, Y. Ponty

Non-redundant sampling



Pick
$$A_{3}$$
, becomes F

Non-redundant Backtrack



Non-redundant Backtrack



Non-redundant Backtrack - Summary

- Forbidden elements increase dependency between local choices
- Need to track which forbidden LMs can be sampled at each stage

Non-redundant Backtrack - Summary

- Forbidden elements increase dependency between local choices
- Need to track which forbidden LMs can be sampled at each stage

Our solution/implementation:

- Tree-like structure to store *Prob*(*A* | *S*, *F*)
- Values used during backtrack
- Updated after generating each LM
- No asymptotic complexity overhead
- Low practical implementation overhead
- Adapted to other algorithms (e.g. RNAsubopt)



Software Implementation

-- -

- C implementation (\approx 3 200 l.), based on Vienna package's RNAlib:
 - Non-redundant sampling
 - Exhaustive enumeration and counting
 - Arbitrary precision arithmetic (MPFR library, L. Fousse *et al*, ACM TOMS 2007)

• Availability:

https://project.inria.fr/rnalands/software/rnanr/

J. Michalik, H. Touzet, Y. Ponty

Results

SV11 (Biebricher and Luce, 1992):

- Metastable state (MS) $\leftarrow \frac{28.5 \text{ kcal.mol}^{-1}}{28.5 \text{ kcal.mol}^{-1}}$ Ground state (MFE)
- $\frac{Prob(MFE)}{Prob(MS)} \approx 10^{12}$
- Hard access to MS-like structures by sampling algorithms
- Exhaustive enumeration in restricted folding space →→ min helix length = 4, ≈ 36000 solutions found

MFE

Speed-up of Non-redundant sampling

T(K) :#redundant structures to obtain K #unique structures
Speed-up: T(K)/K = Avg #times a structure is sampled



Analysis of RNA Artificial Switches

Comparison : RNANR, RNALocopt, RNASLOpt, RNAlocmin & RNAsubopt 250 Bistable sequences: 100nt, 2 structures (one MFE), $\Delta G \leq 5$ kcal.mol⁻¹, ≥ 20 bp distance

Better RNA landscape model = lower switching time Switching time = 1^{st} time t s.t. $Prob_t(MS) < Prob_t(MFE)$





J. Michalik, H. Touzet, Y. Ponty



- RNANR correctly identifies MFE and MS states about 90% of time \rightarrow Expected since most sequences were selected using RNANR
- RNANR most frequently returns lowest switching time

 $\Rightarrow {\tt RNANR} \ {\tt performs} \ {\tt the} \ {\tt better} \ {\tt sampling} \ {\tt to} \ {\tt analyse} \\ {\tt RNA} \ {\tt folding} \ {\tt landscape} \ {\tt than} \ {\tt competitors} \end{cases}$

Conclusion

- RNANR efficiently produces accurate simplified kinetic landscapes
- Non-redundant principles for DP-based sampling, soon in ViennaRNA package
- First step towards the kinetic analysis of larger RNAs

Work in progress (with C. Rovetta)

- Unbiased estimator for structural properties within Boltzmann Ensemble \rightarrow Better convergence
- Compare to exact and simplified kinetics in small systems
- Extending method to pseudoknots
- Integrate within further stochastic sampling methods

Acknowledgments

Thanks to

Advisors

- Yann Ponty
- Hélène Touzet U

Amibio Team

- Mireille Régnier
- Christelle Rovetta
- Afaf Saaidi
- Alice Héliou
- Amélie Héliou



TBI Vienna Team

- Ronny Lorenz
- Gregor Entzian
- Andrea Tanzer
- Maria Waldl
- Ivo Hofacker



Benasque Meeting Organization Team

- Elena Rivas
- Eric Westhof
- Centro de Ciencas Pedro Pascual Team

Montreal McGill Team

- Jérôme Waldispühl
- Matthieu Blanchette
- Carlos Gonzales Oliver
- Roman Sarrazin-Gendron
- Antoine Soulé

🐯 McGill

PhD Funding

• RNALand (ANR)



J. Michalik, H. Touzet, Y. Ponty

Annexes

Results - Proximity of Turner and Combinatorial LMs

Question: Coincidence between Turner and Combinatorial LMs?

Method: Sampling Combinatorial LMs + Gradient descent¹ \rightarrow Final structure, ie Turner Local Minimum

0.547 (0.817)	0.703 (0.757)
avg (std.dev)	avg (std.dev)
$\Delta\Delta G$	Base pair dist.

- More than half Combinatorial LMs (52.4%) are also Turner LMs
- On average, a Combinatorial LM is at ≤ 0.55 kcal.mol $^{-1}$ and 0.7 base pairs to its closest Turner

¹ViennaRNA package, Lorenz *et al*, Algorithms for Molecular Biology 2011

J. Michalik, H. Touzet, Y. Ponty

Non-redundant sampling

Structural Restrictions

- \bullet Space reduction using structural restrictions \rightarrow complexity reduction!
 - \bigwedge Minimum helix length lpha, max #branches within multiloop γ
- \bigwedge Min helix length = 3
- Statistics on RNAStrand (Andronescu et al., 2006)

