

# A study of the folding energy spectrum of RNAs

Jérôme Waldispühl<sup>1 2 3</sup>, Peter Clote<sup>4</sup>

**Keywords:** RNA secondary structure, Boltzmann partition function, locally optimal structures, thermodynamical energy model.

## 1 Introduction.

As the field of *RNAomics* matures, there is an increasing interest in understanding more than just the minimum free energy structure, or low energy ensemble, at thermodynamic equilibrium. A potentially important aspect of the folding process concerns the distribution and free energy of *kinetic traps*, or *locally optimal* secondary structures. Here, a locally optimal secondary structure for a given RNA sequence is a secondary structure, such that no additional base pairs can be added without violation of the definition of secondary structure (e.g. introduction of a pseudoknot). In this paper, we present a new algorithm, which computes for a given RNA sequence  $s = s_1, \dots, s_n$  and all integers  $k$ , the Boltzmann partition function  $Z_k(s) = \sum_{S \in LO(k)} e^{-E(S)/RT}$ , for all locally optimal secondary structures on  $s$  which have exactly  $k$  base pairs ( $E(S)$  is the Turner energy of secondary structure  $S$ ,  $LO(k)$  is the collection of locally optimal secondary structures having exactly  $k$  base pairs,  $R$  is the universal gas constant and  $T$  the absolute temperature).

Graphical display of the computed values  $Z_k(s)$ , as well as that of the *relative density* of states  $\rho_k(s) = Z_k(s) / \sum_k Z_k(s)$ , illustrates an interesting but not yet fully understood difference between functional RNA (e.g. precursor micro-RNA) and random RNA of the same dinucleotide frequency. Given the important functional role of noncoding RNA, as manifested by micro-RNA (which suppress translation of mRNA), riboswitches (which interact with small ligands and up- or down-regulate certain genes), ribozymes (which cleave RNA phosphodiester bonds at specific sites), etc. our current algorithm provides a glimpse into RNA folding landscapes, which may eventually help in the design of RNA for therapeutic purposes – e.g. artificial ribozymes created to cleave HIV and other viral pathogens.

## 2 Methods.

In [1], P. Clote developed an algorithm, which for a given RNA sequence of length  $n$ , computes for each integer  $k$  the number of locally optimal secondary structures having  $k$  base pairs fewer than that of the Nussinov-Jacobson optimal structure. Here, we lift the algorithm of [1] to the Turner energy model [3], and compute for each  $k$  the Boltzmann partition function  $Z_k$  for all locally optimal secondary structures having  $k$  base pairs. In contrast to our earlier algorithm for the Nussinov-Jacobson energy model, which ran in  $\mathcal{O}(n^4)$  time and  $\mathcal{O}(n^3)$  space, the current algorithm for the Turner energy model runs in  $\mathcal{O}(n^6)$  time and  $\mathcal{O}(n^4)$  space, but could be improved to  $\mathcal{O}(n^5)$  time using some algorithmic tricks.

Letting  $Z_k$  denote the Boltzmann partition function for the set of locally optimal secondary structures having  $k$  base pairs, define the relative density of states is defined by

<sup>1</sup>Department of Biology, Higgins 355, Boston College, Chestnut Hill, MA 02467 USA.

<sup>2</sup>Laboratoire d'Informatique (LIX), École Polytechnique, 91128 Palaiseau Cedex, FRANCE.

E-mail: [Jerome.Waldispuhl@polytechnique.edu](mailto:Jerome.Waldispuhl@polytechnique.edu)

<sup>3</sup>LIAFA, Université Denis Diderot, Case 7014, 2 place Jussieu, 75251 Paris Cedex 05, FRANCE.

<sup>4</sup>Department of Biology, Department of computer science (courtesy appt.), Higgins 355, Boston College, Chestnut Hill, MA 02467 USA. E-mail: [cclote@bc.edu](mailto:cclote@bc.edu)

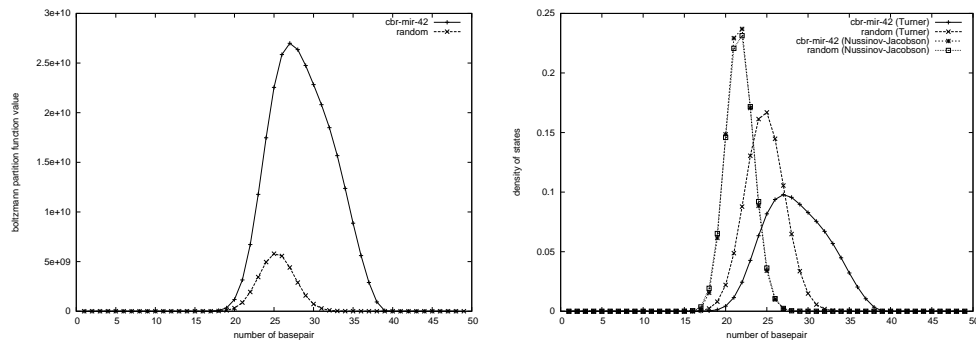


Figure 1: *Caenorhabditis briggsae* miR-42 precursor RNA (cbr-mir-42) and 20 random sequences with the same dinucleotide frequency (here, we show the curve computed, as the average of the 20 random RNA’s curves). Left: Boltzmann partition function. Right: density of states with respect to the Turner energy model, and comparison with the results obtained with the Nussinov-Jacobson model on the same dataset.

$\rho(k) = \frac{Z_k}{\sum_k Z_k}$ . The density of states can be considered as a finite probability distribution, hence we can compute the variation distance, symmetrized relative entropy, and Hodges-Lehmann shift between the density of states of functional RNA and that of random RNA.

### 3 Experiments.

We implemented the previously described algorithm in C, and experimented on a collection of 413 precursor micro-RNA sequences (having length of at most 100 nt.), in each case comparing the Boltzmann partition function and density of states of a fixed miRNA with that of many random RNAs having the same dinucleotide frequency. In order to measure the improvement of our current method over that of [1] (which corresponds to the current method at infinite temperature), we have also computed the Boltzmann partition function (and the relative density of states) with respect to the Nussinov-Jacobson energy model. An example of a typical result is shown in Figure 1, a precursor micro-RNA from *Caenorhabditis briggsae*.

Three major comments should be made. First, the values of the Boltzmann partition function at levels  $k$  are higher (left diagram in Figure 1). Second, the numerical values for the density of states are clearly different, and improved, with respect to the Turner energy model over that for the Nussinov-Jacobson model (right diagram in Figure 1). Third, the density of states distributions are significantly different in case of real RNA sequences and random sequences.

## References

- [1] P. Clote. An efficient algorithm to compute the landscape of locally optimal RNA secondary structures with respect to the Nussinov-Jacobson energy model. *Journal of Computational Biology*, 12(1):83–101, 2005.
- [2] J.S. McCaskill. The equilibrium partition function and base pair binding probabilities for RNA secondary structure. *Biopolymers*, 29:1105–1119, 1990.
- [3] T. Xia, Jr. J. SantaLucia, M.E. Burkard, R. Kierzek, S.J. Schroeder, X. Jiao, C. Cox, and D.H. Turner. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry*, 37:14719–35, 1999.