Algorithmic aspects of RNA locally optimal structures

Bonsai Team – Inria Lille Nord Europe and LIFL

Hélène Touzet – helene.touzet@lifl.fr – CNRS/LIFL

Yann Ponty – yann.ponty@lix.polytechnique.fr – CNRS/LIX, Polytechnique

Scientific context

RiboNucleic Acids (RNAs) are single-stranded macromolecules which can be crudely described as sequences of length ranging from 20 to 3 000 letters, over the four-letters alphabet {A, C, G, U}. Due to its single-stranded nature, any synthesized RNA molecule undergoes a **folding process**, in which connections, also known as base-pairs mediated by hydrogen bonds, are established between its nucleotides. The outcome of RNA folding is a **large variety of structures, or spatial conformation**, which **strongly determine the function of a given RNA** within cellular mechanisms. From a computer science perspective, the structure(s) of an RNA can be acceptably abstracted as a graph, or a tree.

Internship objective

The main goal of this internship is the design and implementation of one (or several) algorithm(s) that would enable the **enumeration and random generation of RNA locally optimal structures**. Such structures act as kinetic traps in the energy landscape, aka local minima of the energy function, and are generally believed to significantly slow-down, or even disrupt, the folding of structured RNAs. Their enumeration represents a mandatory first step towards an **efficient *in silico* analysis of the kinetic behavior of RNA** from its sequence. Previous works [1] have resulted in a polynomial-time algorithm for the enumeration of locally optimal structure in a combinatorial setting based on base-pair maximization. This algorithm will be the starting point of the internship, and it will be completed/extended to capture the complete set of features supported by the – more accurate and realistic – Turner model [2].

Environment

The intern will be hosted by the Bonsai team (Inria Lille Nord Europe et LIFL), and co-supervised by Hélène Touzet and Yann Ponty (CNRS LIX/PIMS Vancouver). It may **start between January and April 2015, and will last from 4 to 6 months**. A compensation (*gratification de stage mensuelle*) may be allocated to the successful candidate.

This work is part of the French/Austrian project RNALands, recently funded by ANR and FWF, whose aim is the design of efficient predictive methods for RNA kinetics. It may be followed by a PhD at LIX (Computer Science Dept of Ecole Polytechnique –

Palaiseau), and include research visits to the other partners of the project, in Lille and Vienna.

Candidate profile

The perfect candidate for this internship is a fifth-year student in Computer Science with a strong background in **algorithms/data structures and/or bioinformatics**, and a real taste for **algorithm design and implementation**. A preliminary experience in C/C++, plus a scripting language of the candidate's choice (Python, Ruby, Perl…), is required. A background in Molecular Biology/Biochemistry, or some measurable level of intellectual curiosity for the subject, will be considered a plus.

To apply, please send:

- a complete resume;
- a cover letter stating your objectives and expectations;
- a copy of your academic record/transcript for the past two years;
- the contact of 2-3 reference;

to Hélène Touzet (helene.touzet@lifl.fr) and Yann Ponty (yann.ponty@lix.polytechnique.fr).

References

[1] Azadeh Saffarian, Mathieu Giraud, Antoine de Monte, and Hélène Touzet. RNA locally optimal secondary structures. *J Comput Biol*, 19(10):1120–1133, Oct 2012.

[2] D. H. Mathews, J. Sabina, M. Zuker, and D. H. Turner. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol*, 288(5):911–940, May 1999.