

PIMS Analytic RNA Combinatorics Workshop (PARC)

April 14 1:30 - 3:00 PM K9509 (Burnaby campus SFU)

Introduction to RNA, Markus Nebel

In this introductory seminar, we will present several ways of how RNA structure is modeled. Starting from a brief explanation of the chemical structure of RNA, we will discuss different but closely related discrete models and sketch their ties to the algorithmic prediction of RNA folding.

April 15 - 16 2014

Simon Fraser University, Burnaby CANADA

TASC II 8500

This two day workshop focuses on the combinatorial aspects of RNA secondary structure.

Speakers:

- Ralf Bundschuh (Ohio State)
- Anne Condon (UBC)
- Christine Heitsch (Georgia Tech)
- Jan Manuch (SFU/ UBC)
- Markus Nebel (TU Kaiserslautern)
- Yann Ponty (CNRS; Polytechnique)
- Cedric Chauve (SFU)

Current RNA bioinformatics strongly relies on discrete abstractions and decomposition schemes for the conformation of nucleic acids. This workshop will bring together scientists from bioinformatics, mathematics and computer science in an attempt to bridge the gap between the enumeration and (asymptotic) analysis of models for RNA structure, and the design of RNA algorithms.

Organizers

Marni Mishna and Yann Ponty

The talks will take place in PIMS seminar room of TASC II 8500

The Introductory talk by Markus Nebel will be in The Shrum Science Center K9509

Abstracts

On the design of exponentially long folding pathways

Anne Condon

It is interesting to understand how DNA or RNA molecules fold, i.e., change their shape or secondary structure, and how to program molecules to fold in useful ways. In particular, DNA computations happen when sets of interacting molecules follow specific folding pathways (sequences of secondary structures). Longer pathways

imply longer and thus potentially more complex computations.

This motivates the question: is it possible to design a set of DNA or RNA molecules that follow a folding pathway of length exponential in the total length of the participating molecules? In this talk we'll show that the answer is yes, for a simple energy model. To design our molecules, we use models and abstraction techniques from the field of molecular programming, namely chemical reaction networks and their realization as DNA strand displacement systems.

The correctness of our design relies on there being a single copy of some participating molecules in solution. We'll also show that the answer to our design question is no if many copies of all molecules are present.

Computing energy barrier of folding pathways

Jan Manuch

Knowledge of energy barriers between pairs of secondary structures for a given RNA molecule is useful, both in understanding RNA function in biological settings and in design of programmed molecular systems. Current heuristics are not guaranteed to find the exact energy barrier, raising the question whether the energy barrier can be calculated efficiently.

We study the computational complexity of a simple formulation of the energy barrier problem posed by Morgan and Higgs, in which each base pair contributes an energy of -1, only base pairs in the initial and final structures may be used on a folding pathway from initial to final structure, each arc from the initial structure is removed exactly once and each arc from the final structure is added exactly once along the pathway, and all the intermediate structures along the pathway are pseudoknot-free. Morgan and Higgs refer to this type of pathways as direct pathways. Note that the length of direct pathway is exactly the number of arcs in the initial structure plus the number of arcs in the terminal structure. The barrier of the pathway is the difference between energies of the highest energy intermediate structure along the pathway and the initial structure. We show that the problem of determining whether there is a direct pathway between two structures with barrier at most k is NP-complete. On the other hand, we show how to solve this problem exactly, using techniques for deconstructing bipartite graphs. The problem is NP-hard and so our algorithm requires exponential time $O(n^{2k+2.5})$ in the worst case but performs quite well empirically on pairs of structures that are hundreds of nucleotides long.

Phase transitions in homopolymer models of RNA

Ralf Bundschuh

The RNA homopolymer model, in which each base interacts with each other base in an identical fashion, has been introduced by de Gennes in 1968 and shown to allow exact calculations of the partition function and thus all thermodynamic quantities. In this talk we will discuss how introducing two different interaction energies between the bases still allows analytical calculations of the asymptotic behaviors of the partition functions and can be used to discuss phase transitions in a number of interesting problems, namely the transition from a native to a molten state, denaturation, aggregation of two RNA molecules, and the role of sequence disorder.

Enumeration of RNA pseudoknot classes

Markus Nebel

RNA molecules are known to have so-called pseudoknots which make structure prediction more complicated -- even for a simple model of free energy, allowing arbitrary pseudoknot conformations to show up makes structure prediction an NP-complete problem. Therefore, various classes of pseudoknots have been introduced over the years, hoping to cover a large portion of native foldings but allowing for efficient prediction algorithms. In this block we will discuss, how the symbolic method of analytic combinatorics

can be extended to easily allow for the enumeration of almost all pseudoknot classes known in literature.

Comparison of RNA secondary structures

Cedric Chauve

In this talk, I will present a family of methods for comparing, mostly pseudoknot-free, RNA secondary structures.

Pseudoknot-free RNA secondary structures can be modelled as ordered rooted trees or, at a higher detail level, arc-annotated sequences. I will describe earlier work on the tree edit distance and tree alignment that show that, unlike for sequences, edit distance and alignments are different for trees. I will also present a surprising result by Herrbach, Denise and Dulucq that shows that the average complexity of aligning tree is quadratic, i.e. of the same order than sequence alignment, while the worst-case time complexity is quartic.

Next I will describe how modelling RNA secondary structures with arc-annotated sequences allows to compare them with improved edit models, and I will present a hierarchy of alignment problems due to Blin and Touzet that enlightens the fundamental difference between edit distance and alignment.

If time allows I will present additional results on alternative models of RNA secondary structure comparisons, such as quotiented trees and chaining problems.

Analytic properties of RNA secondary structures and their representations

Yann Ponty

In this talk, we will recall the basic tools of singularity analysis, and show how they can be used, within a simple four-steps program, to study the asymptotic properties of RNA secondary structures. The whole methodology specializes the symbolic method introduced by Flajolet et al, and allows for multiple extensions, such as the incorporation of weights to model Boltzmann-like distributions. We illustrate this unified framework by deriving various quantities, such as the average distance between the extremities of an RNA, the distribution of occurrences of motifs in a random structure, the average complexity of RNA sampling algorithms, or the asymptotic growth of RNA shapes, some abstract representation of RNA structure. Some of these results, in comparison with experimental data, reveal discrepancies which question the representativity of “pure” combinatorial models (eg homopolymer hypothesis), and will hopefully lead to fruitful discussions during the upcoming afternoon session.

Meanders and RNA Folding

Christine Heitsch

A closed meander of order n is a non-self-intersecting closed curve in the plane which crosses a horizontal line at $2n$ points. Meanders occur in a variety of settings from combinatorial models of polymer folding to the Temperley-Lieb algebra, yet the exact meander enumeration problem remains open. Building on results for plane trees and noncrossing partitions motivated by the biology of RNA folding, we prove that meanders are connected under appropriately defined local move transformations.

The resulting meander graphs have some interesting characteristics and suggest new approaches to the enumeration question. As we will explain, meanders also relate to the challenging biomathematical problem of comparing different possible folds for an RNA sequence.

PARC-Poster.pdf

Workshop Poster