

SEMINAR

Robert Giegerich, Ph.D.
Practical Computer Science
Bielefeld University, Germany

“The Magic of Abstract Shape Analysis”

Abstraction is our major mental aid to master complexity. When we speak about functional structures of RNA, we speak of long hairpins for miRNA precursors, of clover leaf structures for tRNA, of neighbouring hairpins with attenuators, etc., and we do not care about individual base pairs or helix sizes. For programs comparing RNA structures, it has long been suggested to represent larger structures as trees at different level of detail.

RNA structure prediction algorithms, however, are ignorant of abstraction and either deceive us with a single, minimum free energy prediction, or overwhelm us with a plethora of near-optimal structures, most of which are quite similar and therefore redundant.

RNA shape abstraction maps structures to a tree-like domain of shapes, retaining adjacency and nesting of structural features, but disregarding helix lengths. Shape abstraction integrates well with dynamic programming algorithms, and hence it can be applied during structure prediction rather than afterwards. This avoids exponential explosion and can still give us a non-heuristic and complete account of properties of the molecule's folding space. Quite magically, some long hard-studied problems become easy.

In the presentation, I will shortly review the notion of abstract shapes, and explain where its power comes from. I will then provide an overview of several problems that can be approached in a new way:

- (1) Computation of a SMALL set of representative structures of different shapes, complete in a well-defined sense.
- (2) Computation of accumulated shape probabilities, yielding an exact answer to what SFOLD computes via sampling (cf. talk by Ye Ding in this series).
- (3) Comparative prediction of consensus structures, as an alternative to the over-expensive Sankoff Algorithm -- the new algorithm takes linear time in the number of structures, while the Sankoff algorithm is exponential.

I hope to provoke discussions on the biological significance of shape abstractions, and on some issues that are algorithmically more complex than they should be. Tools implementing the above approaches are RNASHAPES and RNACAST, to be found in the "RNA Studio" of the Bielefeld Bioinformatics Server at <http://bibiserv.techfak.uni-bielefeld.de>

Date: Thursday, August 4

Time: 11:00 a.m. - 12:00 p.m.

Location: 1352 Gilman

Presented by the L. H. Baker Center for Bioinformatics and Biological Statistics
and the NIH-NSF BBSI Summer Institute in Bioinformatics and Computational Biology
<http://www.bioinformatics.iastate.edu/seminars/index.html>