

Additional paper for “RNA secondary structure prediction”

Paper Reference

Björn Voß (Voss), Structural analysis of aligned RNAs . Nucleic Acids Research (Oxford Journals). October 2006.

Abstract

In this paper, the author describes the need for structural analysis of different classes of RNAs and discusses RNAshapes, a tool/algorithm which facilitates the prediction of a consensus structure of a class of RNA sequences. The algorithm uses extensions of several techniques from single sequence RNA structure prediction.

The input here is a set of aligned RNA sequences, to which a shape abstraction technique is applied first. The biggest utility of the shape abstraction technique as a first step is that it retains only the nesting and adjacency pattern of helical and unpaired regions, so any exponential growth of the number of sub-optimal regions is curtailed.

Structure prediction (based on an adaptive technique using a grammar describing the search space) is then performed, which accepts user defined inputs too, providing the ability to assign confidence levels to different structures. The grammar used in RNAshapes describes RNA secondary structures without isolated base pairs and handles dangling bases differently. This branching into grammar and algebras facilitates the application of already existing grammars on individual sequences and then averaging it over the entire alignment sequence.

Unpaired bases and gaps that arise due to alignment of different RNA strands are taken into account by considering the thermodynamic energies of various scenarios. Ultimately the consensus structure predicted has the minimum mean free energy (MmFE).

Discussion:

The author has come up with an algorithm that forms a consensual prediction of the given aligned RNA sequences, which is a significant step ahead of the prediction of the structure of a single RNA sequence that was discussed in class. Different (classes of) non-coding RNAs are characterized not by their sequence similarities, but by their structural properties, and under this light, this study could definitely lead to a more generic algorithm.

In my opinion, one of the best aspects of the algorithm was having shape abstraction as the first approach, and then opting for more specific filtering/prediction schemes. This must have been the prime reason for the asymptotic complexity to be significantly less ($O(p^N, N^3, M)$ for an alignment of length N holding M sequences, where p depends on the shape abstraction chosen).

Typically, genetic algorithms that work best have a mix of learning/stochastic and deterministic techniques, with a carefully selected feature vectors (for the training sets/learning algorithms). In the “Structure Prediction” section for instance, the author illustrates a base-pairing parameter, where there is a threshold level which would be user defined. ‘Hardwiring’ the threshold is not considered good practice. Also, in calculating the free energy, the parameters selected to determine the MmFE alignment sequence may/may not be functionally complete.

RNAshapes builds upon existing algorithms for consensual prediction of aligned sequences and manages to keep the asymptotic complexity within bound. It would however be useful to check if self learning schemes could be incorporated into Structure Prediction.