

Tree Decomposition Based Fast Search of RNA Structures Including Pseudoknots in Genomes

Yinglei Song*, Chunmei Liu*, Russell Malmberg†, Fangfang Pan† and Liming Cai*

Abstract

Searching genomes for RNA secondary structures using computational methods has become an important approach to the annotation of non-coding RNAs. In general, the sequence-structure alignment is the most computationally intensive part for such a searching tool. In this paper, we develop a novel graph theoretical approach that can efficiently align a queried sequence to a structural profile with high accuracy. In particular, the consensus structure of an RNA family can be modelled with a conformational graph, the tree width t of which is generally small (e.g., $t = 2$ for stem loops and only a slight increase for pseudoknots). A queried sequence can be preprocessed and an image graph can thus be constructed for it. Based on this framework, sequence-structure alignment can be accomplished by finding in the image graph the maximum valued subgraph that is isomorphic to the conformational graph, which can be computed through dynamic programming on a tree decomposition of the conformational graph in time $O(k^t N^2)$, where k is a small parameter and N is the size of the profiled RNA structure. Experiments showed that this alignment algorithm can achieve the same accuracy as algorithms based on the Covariance model (CM) but is significantly faster. In particular, very accurate searches of tm-RNAs in bacteria genomes and of telomerase RNAs in yeast genomes can be accomplished in days, as opposed to months required by methods based on the Covariance model.

Keywords: Secondary structure profiling, Pseudoknot search, Tree decomposition, Dynamic programming, Covariance model

*Department of Computer Science, University of Georgia, Athens, GA 30602.

†Department of Plant Biology, University of Georgia, Athens, GA 30602.