RNA: Algorithms for structure prediction and gene-finding

P. Clote
Department of Biology
Department of Computer Science (courtesy appt.)
Higgins 355, Boston College, Chestnut Hill, MA 02467, USA
Tel: 617 552 1332, Fax: 617 552 2011, clote@bc.edu.

1 Dec 2004

Abstract

RNA is a current focus of interest in molecular biology, due to post-transcriptional regulatory action of micro-RNA (miRNA) and small interfering RNA (siRNA), which allow geneticists to knock down protein translational products and better understand gene interactions. Genome scanning filters and certain recent RNA noncoding gene-finders have been developed to detect miRNA genes. Using McCaskill’s algorithm for the Boltzmann probability of the low energy ensemble of RNA secondary structures, with a recursive sampling technique, one can now determine potential target regions of mRNA for hybridization with miRNA or siRNA. In another direction, RNA secondary structure plays important roles in retranslation events, such as the incorporation of selenocysteine using the UGA stop codon as well as in ribosomal frameshift slippage events.

This tutorial surveys some of the recent biology of RNA, obtained from analysis of the ribosomal conformation (e.g. non-canonical base pairs such as A-A, dihedral angle preferences), and some important algorithms concerning secondary structure prediction and non-coding gene finders (Zuker’s algorithm, McCaskill’s algorithm for Boltzmann probability, pseudoknot prediction algorithms, Burges miRNA filter, Rivas and Eddy non-coding RNA gene finder for archaeabacteria, analysis of Z-scores, etc.). We additionally include recent results of the author concerning the computation of the landscape of kinetic traps for RNA (to appear in J. Computational Biology), and on asymptotic mean and standard deviation minimum free energy of random RNA (submitted).
Outline

1. Biology of RNA
   (a) Analysis of the structure of the ribosome: non-canonical RNA base pairs, GNRA tetraloops, etc.
   (b) Preferred dihedral angles of RNA nucleotides
   (c) Leontis-Westhof notation for tertiary contacts of RNA, cis/trans Hoogsteen, sugar and Watson-Crick edges, examples
   (d) post-transcriptional gene regulation: miRNA and siRNA, Tuschi’s rules
   (e) retranslation events: incorporation of selenocysteine, ribosomal frameshift and algorithms for determining such retranslation events

Structure prediction
   (a) Nussinov-Jacobson algorithm (we focus initially on this algorithm, since it is very easy to give a good understanding of McCaskill’s algorithm for the partition function, and Ding’s algorithm for sampling from the low energy ensemble with respect to the Nussinov-Jacobson energy model)
   (b) Zuker’s algorithm, McCaskill’s algorithm
   (c) Rivas-Eddy pseudoknot prediction algorithm, Dirks and Pierce refinement for the partition function for pseudoknots
   (d) scanning version of Zuker’s algorithm: lowering updates from $O(n^3)$ to $O(n^2)$ by reusing the energy matrix from the previous computation
   (e) non-coding RNA gene finders: Rivas-Eddy gene finder for archaeabacteria, using $Z$-scores of minimum free energy, Bartel-Burge miRNA gene finder, Coventry-Kleitman-Berger non-coding gene finder
   (f) computing landscape of kinetic traps ($k$-locally optimal secondary structures) of Clote
   (g) computing asymptotic $Z$-scores by precomputing asymptotic mean and standard deviation of minimum free energy of random RNA of fixed dinucleotide frequency (proof of new mathematical result using Kingman’s ergodicity theorem for subadditive stochastic processes – due to Clote, Ferre, Kranakis and Krizanc)
   (h) computing the best face-centered cubic on-lattice representation of RNA Cl’ atoms, to compute threading potentials for RNA (work in progress of Clote)
References


