THE GArna TOOLBOX FOR RNA STRUCTURE ANALYSIS: THE 2006 STATE OF THE ART

Titov I.I.\textsuperscript{1,2}

\begin{itemize}
  \item \textsuperscript{1} Institute of Cytology and Genetics, SB RAS, Novosibirsk, 630090, Russia; \textsuperscript{2} Novosibirsk State University, Novosibirsk, 630090, Russia
  
  e-mail: titov@bionet.nsc.ru
\end{itemize}

Key words: RNA secondary structure, inverse folding, genetic algorithm, miRNA, mRNA, translation

SUMMARY

Motivation: RNA-RNA interactions control numerous processes in cell. There is a continuing demand for computer programs based on analysis of RNA to study RNA structure-function relationships, develop new drugs, analyze genomic sequences.

Results: Here we describe a program package for analysis of RNA-RNA interactions. Most package options are available through the Internet (Titov \textit{et al.}, 2006): simulation of RNA folding, search of RNA fitted to a given secondary structure, calculations of the secondary structure using additional information. GArama provides tools for drawing RNA secondary structure in a fairly quick and easy way. The program package was recently extended with new programs to address the following issues: calculation of RNA secondary structure using energy minimization, design of mRNA 5′UTRs, miRNA gene design, calculation of RNA helix dynamics, design of siRNAs.

Availability: wwwmgs2.bionet.nsc.ru/mgs/systems/garna or on request.

INTRODUCTION

RNA structure defines its function and interactions with other molecules. Algorithms are available via the Internet that makes feasible the calculation of the RNA secondary structure (Hofacker, 2003; Zuker, 2003). Although the thermodynamical algorithms (Hofacker, 2003; Zuker, 2003) remain most wide spread, they can be outperformed by evolutionary programming approach (Wiese, Hendriks, 2006). We have previously reported the GArama server to solve this and inverse problems (Titov \textit{et al.}, 2006), based on genetic algorithm. Modern biotechnology has made practicable the construction of new molecules to expand the functional repertoire of RNA and optimize the processes with its involvement. A recent hot topic of ribonomics is small RNAs. They revealed new horizons for RNA control of cell processes. More and more is known about the role of RNA and there is a growing need in expanding the existing packages by new programs. In this paper we briefly describe the programs for RNA analysis implemented in the GArama package. The programs are divided into three groups:

1) Sequence analysis: search of potentially structured regions in long RNAs; search for miRNA sites that repress translation.

2) Calculation of RNA secondary structure: search of the optimal and suboptimal secondary structures; folding simulation; calculation of oligonucleotide-RNA structure; calculation of the RNA secondary structure using experimental data; calculation of the dynamical properties of RNA helixes.

3) RNA design: RNA design with a given secondary structure; design of the optimal mRNA 5′ UTR; miRNA gene design; design of siRNAs.
The programs for search of potentially structured regions in long RNAs, RNA folding simulation, calculation of oligonucleotide-RNA complex structure, calculation of the RNA secondary structure using experimental data, RNA design with a given secondary structure have been described elsewhere (Titov et al., 2006). The others, performing the search of miRNA sites that repress translation, calculation of the dynamical properties of RNA helixes, design of the optimal mRNA 5′ UTRs, of miRNA genes and of siRNAs to suppress translation have been developed since last report (Titov et al., 2006).

SEQUENCE ANALYSIS

Search of potentially structured regions in long RNAs.

It has described elsewhere how the program performs this function (Titov et al., 2006). The program calculates the nucleotide score that strongly correlates with the secondary structure energy. The matrix for the potential structure and the score profile are displayed at the output. The program predictions can be used for further experimental (Napierala et al., 2005) or computational study of found regions.

Search of miRNA sites that repress translation.

Search is based on the model of the translational arrest by the RNP complex. (for details, see (Titov, Ivanisenko, 2006). The miRNA and mRNA sequences are input, and the program outputs the potential repression sites, if detected.

CALCULATION OF RNA SECONDARY STRUCTURE

Search of the optimal and suboptimal secondary structures.

The program is based on the modified recursive algorithm (Zuker, 2003) and it calculates optimal and suboptimal set of structure in a given energy window.

RNA folding simulation

The program relies on a rapid genetic algorithm and it allows to find the RNA folding intermediates (for details, see (Titov et al., 2002)). The program outputs the energy, the Z-score of the RNA sequence and the RNA structure graph. Drawn RNA secondary structure may be moved and scaled with the mouse.

Calculation of oligonucleotide-RNA complex structure.

This calculation is offered as an option for the preceding program. It is the user’s choice to set the region where RNA interacts with the oligonucleotide. The program can calculate the structure provided that the target region is screened from intramolecule interactions.

Calculation of the dynamical properties of RNA helixes.

The user sets the helix sequence. The program calculates the partition function of the helix and reports the base-pairing probabilities, one position after another.

Calculation of the RNA secondary structure using experimental data.

By using the point-and-click interface, the user sets the paired or unpaired bases. The program minimizes the structure energy with the given constraints ((Titov et al., 2006). After calculation, the user can return to the editor window, modify constraints, and reiterate calculations (Fig. 1). The program may be useful in completing definition of a partly known structure.
RNA DESIGN

RNA design with a given secondary structure.

The program interface is similar to the preceding. The user sets the secondary structure. The program searches the RNA sequence that maximizes the thermodynamic probability of the target structure.

Design of the optimal mRNA 5’ UTR.

As known, various properties of mRNA leader region can affect translational efficiency. These include the competing AUG codons, the nucleotide and secondary structure environment of the start codon and other properties. The user inputs the mRNA leader sequence. The program can stepwise optimize it by nucleotide substitutions as the user wishes. The program can remove false translation starts, optimize the nearest context of AUG codon, get rid of the leader secondary structure, create the stable hairpin by synonymous substitutions for ribosome pausing. The user may set the nucleotide positions that must remain unaltered.

miRNA GENE DESIGN

The miRNA gene is stable hairpin that contains miRNA sequence. The energy of the secondary structure and the probability to reside in hairpin form are the two guideline features in the search of miRNA genes. The candidate gene and miRNA are user’s choice. The program restores miRNA complementarity within the gene and optimizes the thermodynamic characteristics of the gene hairpin.

siRNA design to suppress translation.

The procedure is based on the model of translational arrest by the RNP-mRNA complex. (Titov, Ivanisenko, 2006). The model assumes that a complex of RNP-particles, with each attached to the siRNA-mRNA binding site, suppresses the translation. The specificity of complex formation depends on the interaction of siRNA with the targets and
mRNA structure in-between. The program searches for siRNA, that maximizes the probability of RNP complex formation, and outputs the siRNA candidate list.

CONCLUSION

Our package is developing as a continuous effort of RNA group from Laboratory of Theoretical Genetics of Novosibirsk Institute of Cytology and Genetics. We extended it with a number of new programs since last report; these programs will be installed on web-server GArna (www.mgs2.bionet.nsc.ru/mgs/systems/garna).

ACKNOWLEDGEMENTS

The work was supported by Innovation Project of Federal Agency of Science and Innovation 02.434.11.3004 “Identification of perspective targets for new medicine drugs based on gene network reconstruction”.

REFERENCES