RNALOSS: a web server for RNA locally optimal secondary structures

P. Clote*

Department of Biology, Department of Computer Science (courtesy appointment),
Higgins 355, Boston College, Chestnut Hill, MA 02467, USA

ABSTRACT

RNAomics, analogous to proteomics, concerns aspects of the secondary and tertiary structure, folding pathway, kinetics, comparison, function and regulation of all RNA in a living organism. Given recently discovered roles played by micro RNA, small interfering RNA, riboswitches, ribozymes, etc., it is important to gain insight into the folding process of RNA sequences. We describe the web server RNALOSS, which provides information about the distribution of locally optimal secondary structures, that possibly form kinetic traps in the folding process. The tool RNALOSS may be useful in designing RNA sequences which not only have low folding energy, but whose distribution of locally optimal secondary structures would suggest rapid and robust folding. Website: http://clavius.bc.edu/~clotelab/RNALOSS/.

INTRODUCTION

RNA can play an important functional role in catalysis, e.g. ribozymes are RNA enzymes that cleave RNA phosphodiester bonds at specific sites (1); see (2) for an overview of potential therapeutic applications of ribozymes to cleave mRNAs of oncogenes (*ras* or *bcr-abl*) and viral transcripts (HIV-1), to overcome drug resistance, control arthritis, etc. Additionally, some small molecules can function as drugs acting on RNA. Such is the case for the aminoglycoside and macrolide families of antibiotics, which disrupt RNA translation in prokaryotes by targeting ribosomal (rRNA) (3).

In contrast to mRNA, noncoding RNA (ncRNA) is transcribed from genomic DNA and plays a biologically important role, although it is not translated into protein. Examples of ncRNA include ribozymes, riboswitches, micro RNA, small interfering RNA (4), tRNA, rRNA, etc. Riboswitches have recently been discovered to interact with small ligands and up- or down-regulate certain genes. Breaker and co-workers (5) report the crystal structures of the add A-riboswitch and xpt G-riboswitch aptamer modules, which distinguish between bound adenine and guanine; see (6) for an overview of bacterial riboswitches, and (7) for the structure, as given in the PDB code 1U8D of a guanine-responsive riboswitch with the metabolite hypoxanthine.

RNAomics (8), analogous to proteomics, concerns aspects of the secondary and tertiary structure, folding pathway, kinetics, comparison, function and regulation of all RNA in a living organism. RNAomics requires the application of numerous existent tools, as well as the development of new computational methods. Well-known RNA computational tools include secondary structure prediction web servers `mfold` (9) and `Vienna RNA Package` (10), the `Sfold` web server (11) to sample secondary structures according to the Boltzmann probability distribution, the `tRNAscan-SE` gene finder for tRNA (12), multiple sequence alignment for the statistical detection of RNA secondary structure `MSARI` (13), dynamic programming pairwise sequence-structure alignment `Dynalign` (14), tertiary structure modeling tool `Mc-Sym` (15), etc. Only a few of the many important computational tools for RNA structure prediction, gene finding, alignment, etc. have been listed.

In this paper, we describe the web server `RNALOSS`, based on the algorithm of Clote (16), which computes an aspect of the folding landscape of an RNA nucleotide sequence $s = s_1, \ldots, s_n$. Given s, this algorithm runs in time $O(n^4)$ and space $O(n^3)$, and computes for each k, the number of k-locally optimal secondary structures (explained below). Work by Clote (16) was motivated by the following question, as has been suggested for proteins (17): is it the case that RNA has been under selective pressure to fold rapidly? Using the algorithm of the web server RNALOSS, it appears that structural RNA has a different folding landscape than random RNA of the same dinucleotide frequency; specifically, for small values of k, there appear to be fewer k-locally optimal secondary structures than in random RNA. Related, but distinct work has appeared in (18–21), for discussion see (16).

---

*Tel: +1 617 552 1332; Fax: +1 617 552 2011; Email: clote@bc.edu

METHODS

A secondary structure for an RNA sequence $s = s_1, \ldots, s_n$ is an expression $s = s_1, \ldots, s_n$ involving dot, left and right parenthesis, which is well balanced, such that nucleotides corresponding to matching parentheses are either Watson–Crick complements or GU wobble pairs.

Definition

A secondary structure S on RNA sequence $s = s_1, \ldots, s_n$ is defined to be a set of ordered pairs (i, j), such that $i + 3 < j$ and the following conditions are satisfied.

(i) *Watson–Crick or GU wobble pairs*: If (i, j) belongs to S, then pair (a_i, a_j) must be one of the following canonical base pairs: (A, U), (U, A), (G, C), (C, G), (G, U) and (U, G).

(ii) *Threshold requirement*: If (i, j) belongs to S, then $j - i > 3$; i.e. there must be at least three unpaired bases in a hairpin loop.

(iii) *Non-existence of pseudoknots*: If (i, j) and (k, l) belong to S, then it is not the case that $i < k < j < l$.

(iv) *No base triples*: If (i, j) and (i, k) belong to S, then $j = k$; if (i, j) and (k, j) belong to S, then $i = k$.

A secondary structure is *k*-locally optimal if it has *k* fewer base pairs than the maximum possible number [i.e. than in the Nussinov–Jacobson optimal structure (22,23)], and yet no base pairs can be added without violating the definition of secondary structure (e.g. without introducing a pseudoknot). To illustrate this notion, consider the RNA sequence GGGGCCCCC, which has three as the maximum possible number of base pairs, as given in the structure (((. . .))). There is only one

Figure 1. Screen image of web server RNALOSS. Users may input an RNA sequence consisting of upper or lower case nucleotides A, C, G, T, U, either by uploading a FASTA-format file (A) or by pasting a nucleotide sequence into (B).

structure having 3 bp, so the number of 0-locally optimal secondary structures is 1. On the other hand, there are twelve 1-locally optimal secondary structures and three 2-locally optimal secondary structures. The latter are listed as follows: (i) (...)....(ii) (....)..(iii) ...(.....). The algorithm of (16) uses dynamic programming to compute, for each $i < j$ and each k, the number of k-locally optimal secondary structures on the subsequence $s = s_i, \ldots, s_j$. Additionally, the algorithm must keep track of visible nucleotides and positions, i.e. those external to any base pair [for technical details see (16)].

WEB SERVER

The web server RNALOSS implements a new algorithm, described in (16), running in $O(n^4)$ time and $O(n^3)$ space, which computes for a given RNA sequence $s = s_1, \ldots, s_n$

and all $k \geqslant 0$, the number of k-locally optimal secondary structures for s. An RNA nucleotide sequence may be input by uploading a FASTA-format file or by entering a nucleotide sequence in the blank provided on the web server form. Three tables are returned by RNALOSS: the number of k-locally optimal secondary structures, the relative density of states (i.e. the ratio of number of k-locally optimal structures over the total number of locally optimal structures) and the minimum free energy (mfe) of a sample k-locally optimal secondary structure (for each value of k, RNALOSS computes a single k-locally optimal secondary structure, denoted here as S_k, among the many possible k-locally optimal structures. Since this feature was implemented for debugging purposes, the current version of RNALOSS does not guarantee that S_k has lowest mfe as evaluated by `RNAeval`, over all k-locally optimal secondary structures. For this reason, the energy of

RNALOSS (RNA locally optimal secondary structures)

Output of graph of density of states, as well as number of k-locally optimal structures, sample energy in kcal/mol.

Click here for box plot of density of states

```
Number of nucleotides: 54
Maximum level of suboptimality: 13
Number of base pairs for Nussinov-Jacobson optimal structure:20


Number of locally optimal structures for this RNA
k        Number of k-locally optimal secondary structures

0        318
1        46598
2        935893
3        9.73283E+06
4        6.68754E+07
5        3.30046E+08
6        1.17413E+09
7        2.81655E+09
8        4.06652E+09
9        3.03951E+09
10       9.84908E+08
11       1.08339E+08
12       2.40561E+06
13       1946

Density values
k        density

0        0.000000
1        0.000004
2        0.000074
3        0.000772
4        0.005308
5        0.026194
6        0.093185
7        0.223536
8        0.322740
9        0.241231
10       0.078167
11       0.008598
12       0.000191
13       0.000000
Sum of density values: 1.000000
```

Figure 2. Output of web server RNALOSS on type III hammerhead ribozyme AF170517 from Rfam (24). The web server RNALOSS outputs a table of number of *k-locally optimal secondary structures, for each possible value of k* $\geqslant 0$ (shown here). Additionally, tables for the relative density of states and mfe values of sample *k*-locally optimal secondary structures are displayed in the browser (data not shown).
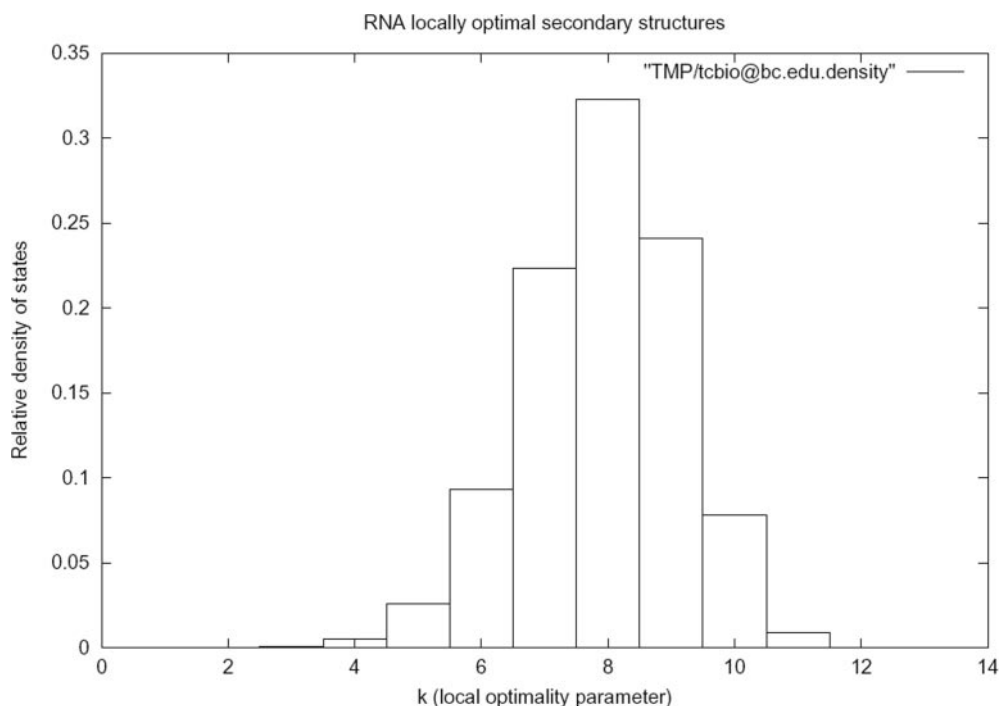
Figure 3. Column graph of relative density of states is obtained by clicking a hot link from the previous screen shot. For each value of $k \geqslant 0$, the ratio of number of k-locally optimal secondary structures over the total number of locally optimal secondary structures is displayed.

sample structures S_k does not necessarily increase monotonically with increasing value of k). For the latter, mfe is computed using `RNAfold` from the Vienna RNA Package http://www.tbi.univie.ac.at/~ivo/RNA/. A screen shot of two of the tables is presented.

Figure 1 displays a screen shot of the RNALOSS web server form. Figure 2 lists the number of k-locally optimal secondary structures as computed by RNALOSS for type III hammerhead ribozyme AF170517 from Rfam (24). Figure 3 presents the relative density of states for k-locally optimal secondary structures for AF170517.

Owing to algorithmic time and space constraints, the RNALOSS web server immediately processes RNA of length at most 60 nt, while for RNA of length 61–100 nt, the results are emailed to the user. Currently, RNALOSS refuses to process any sequence of length >100 nt. Current hardware supporting RNALOSS web server consists of a Beowulf-style cluster comprising 6 Dell 1650, 2 × 1300 MHz Pentium III, 2 GB RAM with 4 Apple XServe, 2 × 1333 MHz G4, 2 GB RAM and finally 6 Dell 1850, 2 × 2800 MHz Xeon EM64T, 2 GB RAM. Interconnect is 1 Gbit Ethernet. Pentium III nodes are running RedHat Linux 9, Xeon EM64T nodes are running WhiteBox Linux 3 and G4 nodes are running MacOS 10.2.8.

DISCUSSION

Upon testing, structurally important RNA, such as selenocysteine insertion sequence elements, precursor mRNAs, type III hammerhead ribozymes and tRNA, all have a markedly smaller number of k-locally optimal structures than that of random RNA of the same dinucleotide frequency, for small and moderate values of k. Since the free energy of k-locally optimal secondary structures is generally closer to that of

the native state for small k, this suggests that structural RNA has been optimized not only to have low folding energy (25), but also to have relatively few potential kinetic traps. This suggests that RNALOSS might be of use in designing RNA sequences for rapid folding.

REFERENCES

1. Doudna,J.A. and Cech,T.R. (2002) The chemical repertoire of natural ribozymes. *Nature*, **418**, 222–228.
2. James,H.A. and Gibson,I. (1998) The therapeutic potential of ribozymes. *Blood*, **91**, 371–381.
3. Sucheck,S., Wong,A.L., Koeller,K.M., Boehr,D.D., Draker,K., Sears,P., Wright,G.D. and Wong,C.-H. (2000) Design of bifunctional antibiotics that target bacterial rRNA and inhibit resistance-causing enzymes. *J. Am. Chem. Soc.*, **122**, 5230–5231.
4. Harborth,J., Elbashir,S.M., Vandenburgh,K., Manninga,H., Scaringe,S.A., Weber,K. and Tuschl,T. (2003) Sequence, chemical, and structural variation of small interfering RNAs and short hairpin RNAs and the effect on mammalian gene silencing. *Antisense Nucleic Acid Drug Dev.*, **13**, 83–106.
5. Serganov,A., Yuan,Y.R., Pikovskaya,O., Polonskaia,A., Malinina,L., Phan,A.T., Hobartner,C., Micura,R., Breaker,R.R. and Patel,D.J. (2004) Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem. Biol.*, **11**, 1729–1741.
6. Barrick,J.E., Corbino,K.A., Winkler,W.C., Nahvi,A., Mandal,M., Collins,J., Lee,M., Roth,A., Sudarsan,N., Jona,I. *et al.* (2004) New RNA

motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc. Natl Acad. Sci. USA*, **101**, 6421–6426.

7. Batey,R.T., Gilbert,S.D. and Montange,R.K. (2004) Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature*, **432**, 411–415.

8. Hofacker,I.L., Priwitzer,B. and Stadler,P.F. (2004) Prediction of locally stable RNA secondary structures for genome-wide surveys. *Bioinformatics*, **20**, 186–190.

9. Zuker,M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.*, **31**, 3406–3415.

10. Hofacker,I.L. (2003) Vienna RNA secondary structure server. *Nucleic Acids Res.*, **31**, 3429–3431.

11. Ding,Y., Chan,C.Y. and Lawrence,C.E. (2004) Sfold web server for statistical folding and rational design of nucleic acids. *Nucleic Acids Res.*, **32**, W135–W141.

12. Lowe,T.M. and Eddy,S.R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.*, **25**, 955–964.

13. Coventry,A., Kleitman,D.J. and Berger,B. (2004) MSARI: multiple sequence alignments for statistical detection of RNA secondary structure. *Proc. Natl Acad. Sci. USA*, **101**, 12102–12107.

14. Mathews,D.H. and Turner,D.H. (2002) Dynalign: an algorithm for finding the secondary structure common to two RNA sequences. *J. Mol. Biol.*, **317**, 191–203.

15. Major,F., Turcotte,M., Gautheret,D., Lapalme,G., Fillion,E. and Cedergren,R. (1991) The combination of symbolic and numerical computation for three-dimensional modeling of RNA. *Science*, **253**, 1225–1260.

16. Clote,P. (2005) An efficient algorithm to compute the landscape of locally optimal RNA secondary structures with respect to the Nussinov–Jacobson energy model. *J. Comput. Biol.*, **1**, 83–101.

17. Šali,A., Shakhnovich,E. and Karplus,M. (1994) How does a protein fold? *Nature*, **369**, 248–251.

18. Cupal,J., Hofacker,I. and Stadler,P. (1996) Dynamic programming algorithm for the density of states of RNA secondary structures. In Hofstädt,R., Lengauer,T., Löffler,M. and Schomburg,D. (eds), *Proceedings of the German Conference on Bioinformatics (Computer Science and Biology)*. Universität Leipzig, Germany, pp. 184–186.

19. Flamm,C., Fontana,W., Hofacker,I.L. and Schuster,P. (2000) RNA folding at elementary step resolution. *RNA*, **6**, 325–338.

20. Flamm,C., Hofacker,I.L., Stadler,P.F. and Wolfinger,M. (2002) Barrier trees of degenerate landscapes. *Z. Phys. Chem.*, **216**, 155–173.

21. Evers,D.J. and Giegerich,R. (2001) Reducing the conformation space in RNA structure prediction. In *Proceedings of the German Conference on Bioinformatics*, pp. 118–124.

22. Nussinov,R. and Jacobson,A.B. (1980) Fast algorithm for predicting the secondary structure of single stranded RNA. *Proc. Natl Acad. Sci. USA*, **77**, 6309–6313.

23. Clote,P. and Backofen,R. (2000) *Computational Molecular Biology: An Introduction*. John Wiley & Sons, NY.

24. Griffiths-Jones,S., Bateman,A., Marshall,M., Khanna,A. and Eddy,S.R. (2003) Rfam: an RNA family database. *Nucleic Acids Res.*, **31**, 439–441.

25. Clote,P., Ferrè,F., Kranakis,E. and Krizanc,D. (2005) Structural RNA has lower folding energy than random RNA of the same dinucleotide frequency. *RNA*, in press.