RNA mutate: RNA secondary structure mutation analysis tool
Alexander Churkin1 and Danny Barash*1,2

Address: 1Department of Computer Science, Ben-Gurion University, 84105 Beer Sheva, Israel and 2Genome Diversity Center, Institute of Evolution, University of Haifa, 31905 Haifa, Israel
Email: Alexander Churkin - churkin@cs.bgu.ac.il; Danny Barash* - dbarash@cs.bgu.ac.il
* Corresponding author

Abstract

Background: RNAMute is an interactive Java application that calculates the secondary structure of all single point mutations, given an RNA sequence, and organizes them into categories according to their similarity with respect to the wild type predicted structure. The secondary structure predictions are performed using the Vienna RNA package. Several alternatives are used for the categorization of single point mutations: Vienna’s RNAdistance based on dot-bracket representation, as well as tree edit distance and second eigenvalue of the Laplacian matrix based on Shapiro’s coarse grain tree graph representation.

Results: Selecting a category in each one of the processed tables lists all single point mutations belonging to that category. Selecting a mutation displays a graphical drawing of the single point mutation and the wild type, and includes basic information such as associated energies, representations and distances. RNAMute can be used successfully with very little previous experience and without choosing any parameter value alongside the initial RNA sequence. The package runs under LINUX operating system.

Conclusion: RNAMute is a user friendly tool that can be used to predict single point mutations leading to conformational rearrangements in the secondary structure of RNAs. In several cases of substantial interest, notably in virology, a point mutation may lead to a loss of important functionality such as the RNA virus replication and translation initiation because of a conformational rearrangement in the secondary structure.

Background

RNAMute is a user friendly computer tool that analyzes point mutations in the secondary structure of RNAs. Initial ideas can be found in [1] and associated works in the late 80’s [2,3]. Since then, much progress has been made in the field RNA secondary structure prediction [4], with the gradual development of sophisticated energy minimization folding prediction packages (most widely used, Zuker’s mfold [5] and the Vienna RNA package [6,7]). The possibility of reliably predicting conformational rearrangements of point mutations in the secondary structure of RNAs has been revisited in [8], suggesting a coarse-grain tree graph representation of the RNA secondary structure [2] and the use of mathematical theorems that relate to eigen-decomposition of the Laplacian matrix [9,10] corresponding to the coarse-grain tree graphs. Both fine-grain and coarse-grain graph representations, including distance measures between the graphs, have been implemented in the Vienna RNA package [6]. We use the Vienna RNA package as the core of RNAMute, attaching to it the muta-
tion prediction procedure described in [8]. To initially test
the approach, experimental results from [11,12] were
taken. Motivation for the use of RNAMute can be found
in the literature [13-16]. These constitute example cases in
which point mutations that affect the functionality of an
RNA molecule cause a conformational rearrangement in
its secondary structure, as explained in detail in the final
Section.

Implementation

Availability

The package can be downloaded from [17]. After down-
loading, extract the file with the commands:

1. >gunzip RNAMute.tar.gz

2. >tar xvf RNAMute.tar

More details on how to run the program are contained in
the ReadMe.html file.

The package content

1. mute_single – performs all possible "single point muta-
tions" in an RNA sequence. The mute_single routine pre-
dicts the secondary structure of the wild type and mutants
using Vienna's RNAfold, then calculates several different
representations and similarity measures between the wild
type and mutants, and finally produces a "result" file from
the results obtained.

2. RNAmute.java – the main routine. Creates a "friendly"
interface for the user. Receives as input a file with an RNA
sequence, runs "mute_single", and generates an HTML file
called "RESULT_TABLE.html" that contains all the proc-
essed data from the "result" file organized in various
tables.

3. calcEig2 – calculates the second smallest eigenvalue of
the Laplacian matrix for each single point mutation.

4. b2Shapiro – converts the full structure from bracket
notation to the weighted coarse grained notation intro-
duced by Bruce Shapiro. This routine uses a function that
is located in the Vienna package's "lib" directory.

5. runRnaMute – similar to RNAmute, but enables the
user to insert the RNA sequence in a text area of the GUI
instead of using a file.

Programs taken from the Vienna RNA package:

1. RNAfold – predicts minimum energy secondary struc-
tures and base pairing probabilities.

2. RNAdistance – calculates the distance between two
RNA secondary structures represented as dot-bracket
strings.

The package also contains the source code for all its com-
ponents.

While the program runs, a new directory called "htmlDir"
will be created. This directory contains all the HTML pages
and all the drawings of the RNA secondary structures that
are being calculated.

Preparation and compilation

RNAMute is currently available on a Linux platform,
therefore all preparations and compilations that will be
mentioned should be performed on a Linux platform with
Java and "GNU CC" compiler installed. RNAMute has all
its components already compiled and may be used with-
out any compilations, although it has some components
written in C that in some architectures may not work. In
such a case, the Vienna RNA package should be down-
loaded from the website [18] and directory "ViennaRNA-
1.4\lib" should be compiled by running the command
"make" in this directory. All files from the directory
"RNAMute\RNAMute_progs" should be copied to "Vien-
naRNA-1.4\Progs" and compiled with "makefile". "Make-
file" that appears in the "ViennaRNA-1.4\Progs" directory
should be overwritten. After the compilation finishes,
files: "b2Shapiro", "calcEig2", "RNAdistance", "RNAfold"
and "mute_single" should be copied from the "Vien-
naRNA-1.4\Progs" directory to the "RNAMute\bin" direc-
tory. All files that are already in the aforementioned
directory should be overwritten. The user should then
make sure that all files in the "RNAMute\bin" directory
are in an executable mode. If not, it is possible to change
their mode by typing the command: >chmod 700
file_name, where file_name is each file from the list
above.

Results and discussion

The input to RNAMute is simply an RNA sequence (see
Figure 1). Subsequently, after pressing the "Start" button,
RNAMute scans all possible single point mutations in that
sequence and computes their folding prediction using
Vienna's RNAfold program. The analysis of point muta-
tions is illustrated in Figures 2, 3 and 4 and will be
described in detail in the manual document file included
in the package. Such an analysis is capable of predicting
conformational rearranging single point mutations, for
example the point mutation that is responsible for switch-
ing between FORM 1 WT RNA and FORM 2 M3 RNA as
described and examined experimentally in [11]. Results
can be observed by pressing the "Result" button. An
HTML page with three tables will appear (Figure 2). For
illustration, we use the IV domain piece that was cut from rRNA of the *Tetrahymena thermophila* [12].

The first table in Figure 2 divides all new structures that were predicted from all point mutations to groups according to their second eigenvalue of the Laplacian matrix [8]. This table also shows how many vertices the structure in each group contains, and the number of structures in each group. In the third column, a group that holds the wild type is marked with "WT", and groups that have the same number of vertices as the WT are marked with "*". The user can click on each value in the first column to view the list of mutations with this value and the specified number of vertices. For example, clicking on eigenvalue 0.381966 (with 5 vertices) will open the table shown in Figure 3. This table contains: (1) mutation's names. (2) distances of the mutations from the WT according to Shapiro's representation for both the mutation and the WT. Mutations in this table are sorted by this column. (3) Minimum Energy (in Kcals/mol) of the secondary structure of mutated sequence. (4) the Shapiro representation of the mutated sequence. Additional information about each mutation can be obtained by clicking on the mutation name. Figure 4 shows the HTML page with additional information for mutation C21G that contains: drawings of RNA secondary structures for the WT sequence and mutated sequence; option to download both drawings in ps format; WT sequence and the mutated sequence; the eigenvalue of the WT secondary structure and of the mutant secondary structure; the WT's free energy and the mutant free energy (in Kcals/mol); Shapiro and dot-bracket representations of both the WT and mutant; distances (according to Shapiro and dot-bracket representations) of mutant from the WT, and the average Shapiro and dot-bracket distances of all mutants.

The second table in Figure 2 divides structures to groups according to their "Dot-bracket distance" from the wild type structure. This distance is calculated between the dot-bracket representations of WT and mutations. The first column contains the distance's ranges that were calculated according to "clustering resolution" for "dot-bracket representation", which is set to 4 by default, and can be changed by the user. Clustering resolution of X means that...
The results for your query:
The wild type sequence:
CGGCUCGUCGCAUCCUCUGGGCGGAAGAAGGUGCUGUUGGCUGUUCGCCAUAAGCGGCACGCGAGCUGG

For more information about a specific eigenValue, press the required link

| Second eigenvalue | Number of vertices | wild type | Frequency |
|-------------------|-------------------|-----------|-----------|
| 0.381966          | 6                 |           | 16        |
| 0.381966          | 5                 | *         | 2         |
| 0.518806          | 5                 | WT        | 193       |
| 1.000000          | 4                 |           | 18        |

For more information about the mutations in a specific distance range, press the required link

| Dot-Bracket Distance range | Frequency |
|---------------------------|-----------|
| 0.0 - 12.0                | 202       |
| 16.0 - 26.0               | 18        |
| 30.0 - 34.0               | 7         |
| 38.0 - 38.0               | 2         |

* the average of Dot-Bracket Distances is: 6.0614
Clustering Resolution : 4

| Shapiro Distances range | Frequency |
|-------------------------|-----------|
| 0.0 - 45.0              | 227       |
| 60.0 - 62.0             | 2         |

* the average of Shapiro's Distances is: 7.6447
Clustering Resolution : 6

Figure 2
Categorization Tables. Main HTML page for the categorization of single point mutations according to the mutants secondary structure similarities relative to the wildtype structure.

distances are sorted in each group and if there are two distances such that the difference between them is less than X, these distances are in the same group.
The user can click on a specific distance range in the first column to observe the list of mutations with a distance in this range. For example, distance range of 38.0-38.0 has a similar table as in Figure 3 and has only 2 mutations. This

The mutations with eigenvalue = 0.381966 and 5 vertices
(The mutations are sorted according to their Shapiro-Distance)
* the average of Shapiro's Distances is: 7.6447

| Mutation | Shapiro-Distance | Minimum Energy Kcal/mol | Shapiro representation |
|----------|------------------|-------------------------|------------------------|
| C45A     | 60.0             | -35.00                  | (((((((((H3)87)120)33(B4)89)1)2)83(B4)88)E1)R) |
| C21G     | 62.0             | -33.02                  | (((((((((H3)85)120)1)1)5(B1)81)81)87)1(3)83(B4)88)E1)R) |

Figure 3
Mutations Descriptor Data. Mutations descriptor HTML page. For each category, the single point mutations belonging to each category are listed along with their minimal free energy and string representation.
distance range is interesting to explore because it contains structures of mutations with a relatively large dot-bracket distance from WT. Additional information about each mutation in each table can be obtained by pressing on the mutation name, such as in Figure 3 and the information page that will be obtained as depicted in Figure 4. In our case these are the same two mutations as were obtained from the first table (eigenvalue 0.381966) and these are the only mutations in the run that break one of two hairpins and linearize the structure.

**Information About This Mutation:**

- **Wild-Type Sequence:**
  
  CGGCUCGUGCUCAUCCUGGGCCUGAAGAAGUCCCAAAGGUUGGTCUGUUCGCCCAUUAAAGCCGGCACGCGACCGUGG

- **Mutation Sequence:**
  
  CGGCUCGUGCUCAUCCUGGGCCUGAAGAAGUCCCAAAGGUUGGTCUGUUCGCCCAUUAAAGCCGGCACGCGACCGUGG

- **Wild-Type EigenValue:** 0.518806
- **Mutation EigenValue:** 0.381966

- **Wild-Type Free Energy:** -38.00 Kcal/mol
- **Mutation Free Energy:** -33.02 Kcal/mol

- **Wild-Type Shapiro Representation:** ((((((H6)S6)B1)S3)(H5)S5)(M9)S3)B4(S8)E1)R
- **Mutation Shapiro Representation:** ((((((H3)S5)B1)S2)E4)S7)(E13)S3)B4(S8)E1)R
- **Mutation Shapiro Distance:** 62.0 (avg: 7.6447)

- **Wild-Type Dot-Bracket Representation:** ((((((...(((......))))))...(((......))))...))...))
- **Mutation Dot-Bracket Representation:** ((((((...(((...(((((......)))))))))...)))...))))
- **Mutation Dot-Bracket Distance:** 38.0 (avg: 6.0614)

**Figure 4**

**Mutant vs. Wildtype Structure and Energy Information.** For each single point mutation, relevant secondary structure and energy information is listed along with a graphical drawing for both the mutant and the wildtype. This allows a direct comparison between the corresponding mutant and the wildtype structure.
Figure 5
Single Point Mutation Prediction in the 5'UTR of HCV by RNAMute. A successful prediction by RNAMute, illustrating its potential capability to detect biologically meaningful findings. The G235A point mutation (corresponding to G95A using our indexing scheme) in the 5'UTR of HCV [16] is predicted by RNAMute to cause a conformational rearrangement. In turn, it is reported to display a dramatic reduction in translation initiation. However, in that reference [16] based on simple base pairing considerations, it was stated that this mutation alters only the primary sequence. With the availability of RNAMute, alterations in the secondary structure can easily be detected.
The third table in Figure 1 is similar to the second table but it groups structures according to their Shapiro distance which is obtained from the Shapiro representation of the WT and mutation’s structure. It is possible to see that the third table also groups two mutations with a relatively large distance to a separate category, and these two mutations are exactly the same mutations that were found in “Eigenvalue table” and “Dot-bracket table”.

From the illustrated example we can conclude that the RNAMute package was able to find mutations that change the secondary structure of the wildtype and it divided these mutations into separate categories in all tables. In the first table these mutations fall to the category with specific second smallest eigenvalue of the Laplacian matrix corresponding to the coarse-grain tree graph representation; in the second and the third tables these mutations fall into categories with largest distances.

Conclusion
In examining its biological relevance, RNAMute can be used in predictions and analyses related to mutagenesis experiments. For example, in [13] it was shown that individual point mutations are capable of inactivating spectinomycin resistance in Escherichia coli and secondary structure predictions displayed conformational rearrangements. Moreover, in examples where the sequences examined contain less than 100 nt, virologists have shown interest in computerized predictions of mutations that disrupt the stable stem-loop structure that characterizes Hepatitis C Virus (HCV) [14-16]. Such structural changes may lead to alterations in virus replication [14,15] or translation initiation [16]. In the latter reference [16], the single point mutations A172G, G229A, and G235A were found to display a dramatic reduction in translation initiation in site-specific mutagenesis experiments affecting the stem-loop IIIc. While it was obvious that A172G and G229A disrupt the base pairing required to form the structures in and around stem-loop IIIc, G235A was assumed to only alter the primary sequence since no obvious Watson-Crick base pairing modifications appear at first glance. However, using RNAMute, G235A can be found to disrupt the important stem-loop structure as well (Figure 5). where G95A according to our indexing scheme corresponds to G235A in the indexing scheme used in [16]. In Figure 5, we only used a segment of the HCV RNA as our initial sequence to RNAMute after verifying that the wildtype of the segment is accurately predicted by mfold and Vienna’s RNAfold. Thus, with the public availability of RNAMute, computational mutation predictions that are needed to detect novel functional biological findings can be improved.

Availability and requirements
Project name: RNAMute
Project home page: http://www.cs.bgu.ac.il/~RNAMute
Operating system(s): web access: not applicable, stand-alone: LINUX
Programming language: C, Java
Other requirements: stand-alone:Java 1.4.0 or higher, GNU CC compiler
License: None
Any restrictions to use by non-academics: None

Authors’ contributions
DB conceived the study, coordinated and participated in software design and drafted the manuscript. AC worked on software design, carried out development and implementation, and participated in drafting the manuscript.

Acknowledgements
The authors would like to thank Ornit Elgad and Oshrat Petgho for their devoted assistance at the initial stages of the RNAMute package development. The research was supported by the Lynn and William Frankel Center for Computer Sciences and a grant from the Israel USA binational foundation BSF 2003291.

References
1. Margalit H, Shapiro BA, Oppenheim AB, Maizel JV: Detection of common motifs in RNA secondary structures. Nucleic Acids Res 1989, 17:4829-4845.
2. Shapiro BA: An algorithm for comparing multiple RNA secondary structures. Comput Appl Biosci 1988, 14:387-393.
3. Le SY, Nussinov R, Maizel JV: Tree graphs of RNA secondary structures and their comparisons. Comput Appl Biosci 1989, 5:461-473.
4. Zuker M: Calculating nucleic acid secondary structure. Curr Opin Struct Biol 2000, 10(3):303-310.
5. Zuker M: Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 2003, 31:3406-3415.
6. Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P: Fast folding and comparison of RNA secondary structures. Monatsh Chem 1994, 125:167-188.
7. Hofacker IL: Vienna RNA secondary structure server. Nucleic Acids Res 2003, 31:3429-3431.
8. Barash D: Deleterious mutation prediction in the secondary structure of RNAs. Nucleic Acids Res 2003, 31(22):e578-6584.
9. Fiedler M: Algebraic connectivity of graphs. Czechoslovak Math J 1973, 23:298-305.
10. Merris R: Characteristic vertices of trees. Lin Mult Alg 1987, 22:115-131.
11. LeCuyer KA, Crothers DM: Kinetics of an RNA molecular switch. Proc Natl Acad Sci 1994, 91:3373-3377.
12. Yusupov MM, Yusupova GZ, Baucom A, Lieberman K, Earnest TN, Cate JH, Noller HF: Crystal structure of the ribosome at 5.3 A resolution. Science 2001, 292(5518):883-96.
13. Zimmerman JM, Maher LJ 3: In vivo selection of spectinomycin-binding RNAs. Nucleic Acids Res 2002, 30(24):5425-5435.
14. Smith DB, Simmonds P: Characteristics of nucleotide substitution in the Hepatitis C Virus genome: constraints on the sequence change in coding regions at both ends of the genome. J Mol Evol 1997, 45:238-246.
15. You S, Stump DD, Branch AD, Rice CM: A cis-Acting replication element in the sequence encoding the NS5B RNA-dependent RNA polymerase is required for Hepatitis C Virus RNA replication. J Virol 2004, 78:1352-1366.
16. Tang S, Collier AJ, Elliott RM: Alterations to both the primary and predicted secondary structure of stem-loop IIIc of the Hepatitis C Virus 1b 5' untranslated region (5' UTR) lead to mutants severely defective in translation which cannot be complemented in trans by the wild-type 5' UTR sequence. J Virol 1999, 73:2359-2364.
17. RNAMute [http://www.cs.bgu.ac.il/~RNAMute]
18. Vienna RNA package [http://www.rna.tbi.univie.ac.at/~ivo/RNA]