

Mutational analysis in RNAs: comparing programs for RNA deleterious mutation prediction

Danny Barash and Alexander Churkin

Submitted: 20th May 2010; Received (in revised form): 8th July 2010

Abstract

Programs for RNA mutational analysis that are structure-based and rely on secondary structure prediction have been developed and expanded in the past several years. They can be used for a variety of purposes, such as in suggesting point mutations that will alter RNA virus replication or translation initiation, investigating the effect of deleterious and compensatory mutations in allosteric ribozymes and riboswitches, computing an optimal path of mutations to get from one ribozyme fold to another, or analyzing regulatory RNA sequences by their mutational profile. This review describes three different freeware programs (RNAMute, RDMAS and RNAmutants) that have been developed for such purposes. RNAMute and RDMAS in principle perform energy minimization prediction by available software such as RNAfold from the Vienna RNA package or Zuker's Mfold, while RNAmutants provides an efficient method using essential ingredients from energy minimization prediction. Both RNAMute in its extended version that uses RNAsubopt from the Vienna RNA package and the RNAmutants software are able to predict multiple-point mutations using developed methodologies, while RDMAS is currently restricted to single-point mutations. The strength of RNAMute in its extended version is the ability to predict a small number of point mutations in an accurate manner. RNAmutants is well fit for large scale simulations involving the calculation of all k -mutants, where k can be a large integer number, of a given RNA sequence.

Keywords: RNA mutational analysis; RNA secondary structure prediction

INTRODUCTION

In recent years, the importance of RNA mutational analysis and the amount of efforts devoted to it has been steadily growing. New discoveries in non-coding regulatory RNAs, as well as notable advances in the understanding of RNA viruses, have led to an increasing number of mutagenesis experiments and in turn to the development of programs that can computationally predict and analyze the effect of unique point mutations on the structure and function of RNAs. In a variety of cases of biological importance, ranging from hepatitis C Virus (HCV) replication [1] and translation initiation [2] to bacterial resistance against antibiotics [3] or to

the function mechanism of spliced leader RNA [4], it was shown that point mutations causing a conformational rearrangement in the RNA secondary structure may bring about to a complete change in the function. To assist in the computational predictions and analyses of such cases that are anticipated to significantly increase in the future, various programs have been developed. This review compares between the three main programs developed for this purpose, namely RNAMute [5, 6] and RDMAS [7] and RNAmutants [8, 9] with their respective approaches. All of them rely on energy minimization, either by a direct utilization of available software for RNA folding prediction that use thermodynamic parameters

Corresponding author. Danny Barash, Department of Computer Science, Ben-Gurion University, 84105 Beer-Shera, Israel.
E-mail: dbarash@cs.bgu.ac.il

Danny Barash is a computational scientist and the advisor and coordinator for bioinformatics in the computer science department at Ben-Gurion University. His research interests include computational biology, RNA structure predictions, and scientific computing. **Alexander Churkin** is a computational scientist pursuing his PhD studies in the bioinformatics track within the computer science department at Ben-Gurion University. His research interests include computational biology, RNA structure predictions, and scientific computing.

[10, 11], an approach that began in the work of [12], or by independently applying the basic recursions that [10, 11] contain with relevant extensions.

When attempting to provide solutions to the problem of RNA deleterious mutation prediction, a logical step is to concentrate on the detection of conformational rearrangements in the RNA secondary structure. It is well known that some RNA secondary structure motifs, such as stem-loops, are important for certain functions and if a conformational rearrangement occurs then their functionality can be disrupted. Therefore, being able to reliably predict the secondary structures of a given wild-type sequence and that of its neighboring mutants becomes useful for deleterious mutation prediction. Alternatively, even if the secondary structure is maintained, the free energy landscape of mutants can change, making certain structures less populated as compared to others, which may even have disease-related consequences in certain systems. All of these scenarios are highly relevant to the problem of RNA deleterious mutation prediction and can be simulated by available software such as the Vienna RNA package or Zuker's Mfold and their extensions that in essence contain a thermodynamic model. For example, making certain structures less populated than others can be simulated by calculating suboptimal solutions to the RNA folding prediction by energy minimization. The exact strategy to be taken within the framework of this model is what distinguishes between the various programs described in this review, with the possibility to improve existing venues or to develop new ones.

Historically, the pioneering work of Woese in the 1970s should be addressed when reviewing the topic of RNA mutations. In his models of ribosomal RNA structure (e.g. [13]) that resulted in a breakthrough in phylogenetic taxonomy, each helix in the secondary structure model has been substantiated by examples wherein significant base changes in its primary structure have occurred without violating the proposed base-pairing constraint. In this way, an important principle was established that correlated nucleotide changes (mutations) can relate to RNA secondary structure in useful ways for structure determination. This work, most of which was done manually or assisted by very simple computer programs, later continued with the Noller–Woese–Gutell comparative structure models for ribosomal RNAs and the development of the RNA comparative analysis approach [14]. During the late 1980s, Shapiro and

coworkers were instrumental in computationally examining the effect of single-point mutations on RNA secondary structure using an RNA tree-graph representation and tree comparisons [15–18]. For RNA secondary structure prediction, in some of these works the Nussinov–Jacobson algorithm was used [19], emanating from the original dynamic programming formulation by Nussinov *et al.* [20] that contributed substantially and is widely taught when introducing all current-day dynamic folding algorithms for secondary structure prediction. The work in [16] examined the prediction of single-point mutations in the L11 mRNA of *Escherichia coli*. Since then, the energy minimization field has advanced considerably both in terms of computational methods (e.g. [21–23]) and biochemical empirical parameters [24]. This has led to much better prediction capabilities at the beginning of the century, and thereafter the problem of RNA deleterious mutation prediction was revived [25]. In [25], the problem was revisited in detail with riboswitch examples. Besides a potential application for the rational design of riboswitches, a biological motivation behind such work is also to investigate the effects of deleterious and compensatory mutations in an RNA allosteric system, in a similar way to what was done beforehand in protein systems such as GroEL [26]. In this context, the method of correlated mutation analysis of Horowitz and coworkers [26, 27] should be mentioned as a complementary approach. Noivirt and Horowitz adapted their correlated mutation analysis to nucleic acids and thereafter were able to discover independently the unique effect of a deleterious mutation that was singled out in [25] by using energy minimization, thus supporting this suggestion from [25] using a different method. Indeed, biochemical experiments performed around this period and reviewed in [28] were able to verify the bulk of the predictions mentioned in [25], demonstrating the success of the approach that led to RNAMute. Next, the development of the RNAMute program has begun, first using an example from bacterial resistance to spectinomycin [3] by showing the concept with a specific software application [29], then as a general software tool for RNA single-point mutation analysis [5] that is structure based and relies on routines from the Vienna RNA package [10] in its core. The first publication came out in [30]. At around this time, Clote and coworkers developed independently an approach that computes the energy landscape of k -point mutants of an RNA molecule [31], which

led to the development of RNAMutants. In the following year, a webserver was put up that is conceptually close to the RNAMute Java tool for single-point mutations but with some different features that will be described in the sequel, called RDMAS [7]. Later, the methodology articles describing in detail the RNAMute approach to handle multiple-point mutations efficiently [6] and the RNAMutants approach to compute all k -point mutants [8] appeared, following with the appearance of the RNAMutants webserver [9]. A tabular overview providing basic information about all three programs is available in Table 1.

DETAILS OF USE

Installing the different programs or accessing the webservers is not a difficult task. They all contain a ReadMe file or a manual that is easy to follow with no prior knowledge assumed.

RNAMute can be downloaded in its latest version that handles multiple-point mutations from <http://www.cs.bgu.ac.il/~RNAMute/Efficient> (note the final word ‘Efficient’ in its name), available in the Department of Computer Science at Ben-Gurion University. It is possible to successfully run single-point mutations from this website as well, although the original version that only handles single-point mutations available at <http://www.cs.bgu.ac.il/~RNAMute> contains some more features for this specific task, relative to the ‘Efficient’ website. The zipped file called MultiRNAMute (389 kB) is currently available for a Linux platform with Java and ‘GNU CC’ compiler installed. It was checked that its GUI option works well on various Linux-based operating systems including Fedora 12. In its core, it uses both RNAfold and RNAsubopt from the Vienna RNA package, and as instructed it is also possible to download their latest version from the

Table 1: A tabular overview with some basic information about RDMAS, RNAMute and RNAMutants

Programs	Properties			
	Handles single-point mutations	Handles multiple-point mutations (k -point mutants)	Uses Vienna's RNAfold and RNAsubopt (or Mfold)	Uses thermodynamic parameters
RDMAS	Yes	No	Yes	Yes
RNAMute	Yes	Yes	Yes	Yes
RNAMutants	Yes	Yes	No	Yes

Vienna RNA package website. In the ReadMe file, the user can find detailed instructions on preparation and compilation.

RDMAS is a webserver that can be easily accessed at <http://biotech.bmi.ac.cn/rdmas/index.php>. It is maintained at the Beijing Institute of Radiation Medicine. It is freely available at no charge to the scientific community for non-commercial applications, but licensing is required for commercial applications, as explained in the website. There is a manual available that explains how to use the webserver in detail.

RNAMutants is available both as a webserver and for download. For download, zipped files (around 2 MB) for various Linux platforms and Mac OS-X are available for non-profit research and educational purposes at <http://rnamutants.csail.mit.edu> of the MIT Computer Science and Artificial Intelligence Laboratory. Commercial organizations can request a for-profit license by contacting the address mentioned. The webserver can be accessed at <http://bioinformatics.bc.edu/clotelab/RNAMutants/index.spy>, residing in Boston College at the Clote Bioinformatics Laboratory. A manual with detailed examples is available for the webserver and additional Python code that may be useful for the creation of mutational profiles in conjunction with RNAMutants can be found at the same website.

RNAMute in detail

RNAMute [5, 6] is an interactive Java application that provides a user friendly interface for analyzing the effects of point mutations on RNAs secondary structure. If the user is interested only in single-point mutations, it is recommended to download the application from <http://www.cs.bgu.ac.il/~RNAMute>, in which case all single-point mutations will be examined with no efficient considerations. If the user is also interested in multiple-point mutations, it is recommended to download the application from <http://www.cs.bgu.ac.il/~RNAMute/Efficient>, which uses an efficient method based on Vienna's RNAsubopt [23]. The basic strategy behind the efficient method is that initially, RNAsubopt is applied to compute all suboptimal solutions of the given sequence, and then after an appropriate filtering step only the mutations that stabilize the suboptimal solutions and destabilize the optimal one are considered as candidates for being deleterious. The complete method is described in [6]. In both cases mentioned above, there are

detailed instructions for the user in a ReadMe file available in the corresponding websites on how to download, compile, and use the program in a Linux environment on a standard PC. Here it is described step by step how the user can analyze multiple-point mutations using the 5BSL3.2 of a subgenomic HCV replicon taken from [1] as an example sequence (See Figure 1).

First, the user enters <http://www.cs.bgu.ac.il/~RNAMute/Efficient>, downloads the application, extracts the files (easy to follow instructions are available in the ReadMe file if necessary) and issues the command 'java Main' following the prompt in the Linux Shell, from the MultiRNAMute directory. Second, the user inputs the following data and parameters into the GUI: the sequence AGC GGG GGA GAC AUA UAU CAC AGC CUG UCU CGU GCC CGA CCC CGC,

dist1 = 20, dist2 = 20, e range = 16, mutations = 3, distance = Hamming, method = Fast, stabilizing and destabilizing. Third, the user presses the 'Start' button and after waiting for ~6 min, the table with 3-point mutations that have a structure with Hamming distance >20 from the wild-type structure appears. The results are shown in Figure 2. After choosing the highlighted mutation in Figure 2 by a single click, a new window with additional information will be opened that includes drawings of the wild-type and mutation structure with some additional information, as displayed in Figure 3.

When choosing the distance parameter, it should be noted that in the efficient version for multiple-point mutations the user can select between the base pair distance and the Hamming distance whereas in the original version for single-point mutations no such choice was needed to be specified; the output

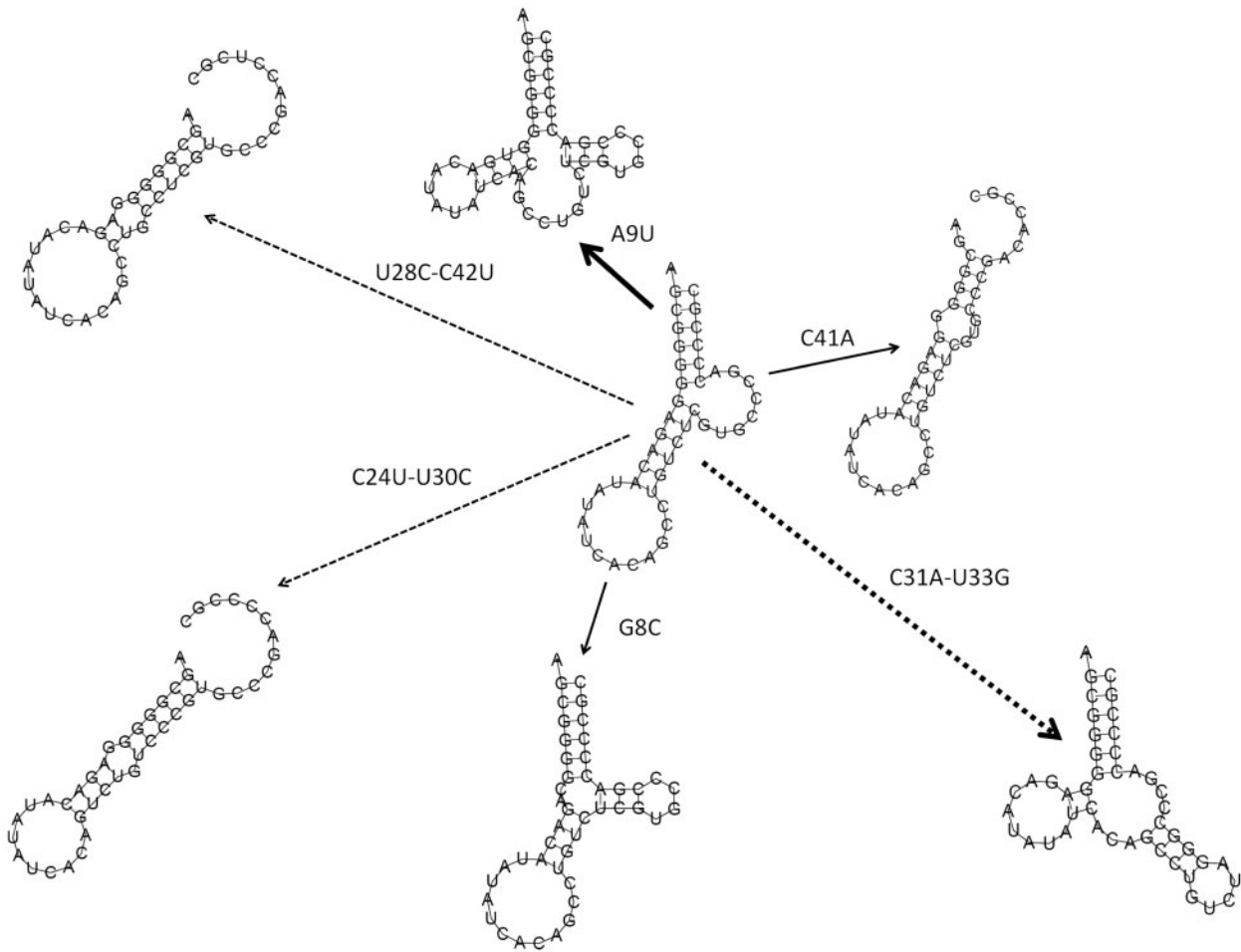


Figure 1: Illustration of RNA structural mutation analysis and conformational rearranging mutations. Starting from the wild-type sequence and secondary structure for the 5BSL3.2 of a subgenomic HCV replicon [1] in the middle, various single-point and double-point mutations in the sequence are shown, resulting in various predicted structural changes.

Mutation Name	Energy (kcal/...)	Distance	Dot-bracket representation
A8U-G43U-C44A	-10.0	22	((((((.....)))))).....
A8U-G43U-C44U	-10.0	22	((((((.....)))))).....
A8U-U17A-C42A	-8.4	22	((((((.....)))))).....
C11A-C24A-C36U	-15.9	22	((((((.....)))))).....
C11A-U32G-C36U	-17.8	22	((((((.....)))))).....
C11G-U32G-C36U	-18.6	22	((((((.....)))))).....
C11U-C24A-C36U	-17.0	22	((((((.....)))))).....
C11U-C30G-C40G	-11.9	22	((((((.....)))))).....
C11U-C30U-C40A	-9.2	22	((((((.....)))))).....
C11U-C30U-C41G	-11.7	22	((((((.....)))))).....
C11U-U32G-C36U	-18.9	22	((((((.....)))))).....
C20G-A21G-C24A	-23.0	22	((((((.....)))))).....
C20G-A21G-U32G	-24.9	22	((((((.....)))))).....
C20G-C24A-C35U	-20.9	22	((((((.....)))))).....
C20G-U27A-C40G	-10.4	22	((((((.....)))))).....
C20G-U32G-C35U	-22.8	22	((((((.....)))))).....
C24A-C28A-C36U	-15.7	22	((((((.....)))))).....
C24A-C28A-G43A	-11.7	22	((((((.....)))))).....
C24A-C28U-C36U	-15.7	22	((((((.....)))))).....
C24A-C35U-C36G	-20.2	22	((((((.....)))))).....
C24A-U27A-C36U	-15.7	22	((((((.....)))))).....
C24A-U27C-C36U	-15.7	22	((((((.....)))))).....
C24A-U29A-C36U	-15.7	22	((((((.....)))))).....
C24A-U29C-C36U	-15.7	22	((((((.....)))))).....
C24A-U29G-C36U	-15.7	22	((((((.....)))))).....
C24G-U27A-C40G	-10.5	22	((((((.....)))))).....
C28A-C40A-G43A	-6.7	22	((((((.....)))))).....
C28A-C40A-G43U	-6.7	22	((((((.....)))))).....
C28A-C40U-C42A	-6.7	22	((((((.....)))))).....
C28A-C40U-G43A	-6.7	22	((((((.....)))))).....
C28A-C40U-G43U	-6.7	22	((((((.....)))))).....

Figure 2: Mutation group list screen as a result of running RNAMute for the case of 3-point mutations for the 5BSL3.2 wild-type [1] shown in the middle of Figure 1. A run for the case of 2-point mutations that succeeded to detect a deleterious 2-point mutation described in [1] can be found in [6].

contained the tree-edit distance computed by Vienna's RNAdistance and some other similarity measures. The difference stems from the expensive computational step of filtering suboptimal solutions in the case of multiple-point mutations (details are available in [6]). Because of that step, the fastest distance methods were implemented. Future extensions, in addition to implementing a tree-edit distance that is more computationally demanding for handling multiple-point mutations, but perhaps can be applied to the categorization of mutations separately from the suboptimal filtering expensive step, may include RNAforester [32] for the tree comparisons and RNASHapes [33] for the coarse-grain representations. Currently, for multiple-point mutations, only the base pair distance and the Hamming distance are available as options for the user, because they are both the fastest possible to compute. In performing mutational analysis by filtering and categorization, they both give very similar results, and therefore picking either one is legitimate. More explanation about their subtleties is available in the methodology paper for the efficient RNAMute [6].

Except for the input sequence that the user needs to insert, all other parameters have default values. In the parameters, however, there is a variety of options

that the user can decide upon, such as number of mutations to run the analysis on, parameters that deal with the amount of computations versus accuracy such as dist1, dist2, e range, and method type that are explained in detail in the ReadMe file and in [6], and distance type to be used (currently, a choice between Hamming distance or base pair distance as discussed above) when calculating distances between RNA secondary structures. Briefly stated, dist1, dist2 and e range are parameters that control how much filtering is performed on the suboptimal solutions that are obtained as a consequence of running RNASubopt [23]. The parameters dist1 and dist2 are the distances used for filtering the suboptimal solutions that are close to the optimal solution and for filtering the suboptimal solutions that are close to each other, respectively. The parameter e range is the energy parameter used with RNASubopt when calculating suboptimal structures within range kcal/mol of the minimum free energy.

In our example for the case of 3-point mutations described above and depicted in Figures 2 and 3, no biochemical experiments were reported yet that can validate or disprove the list of conformational rearranging mutations suggested by RNAMute, but in the example that can be found in the ReadMe file and also described in [6] for the case of 2-point

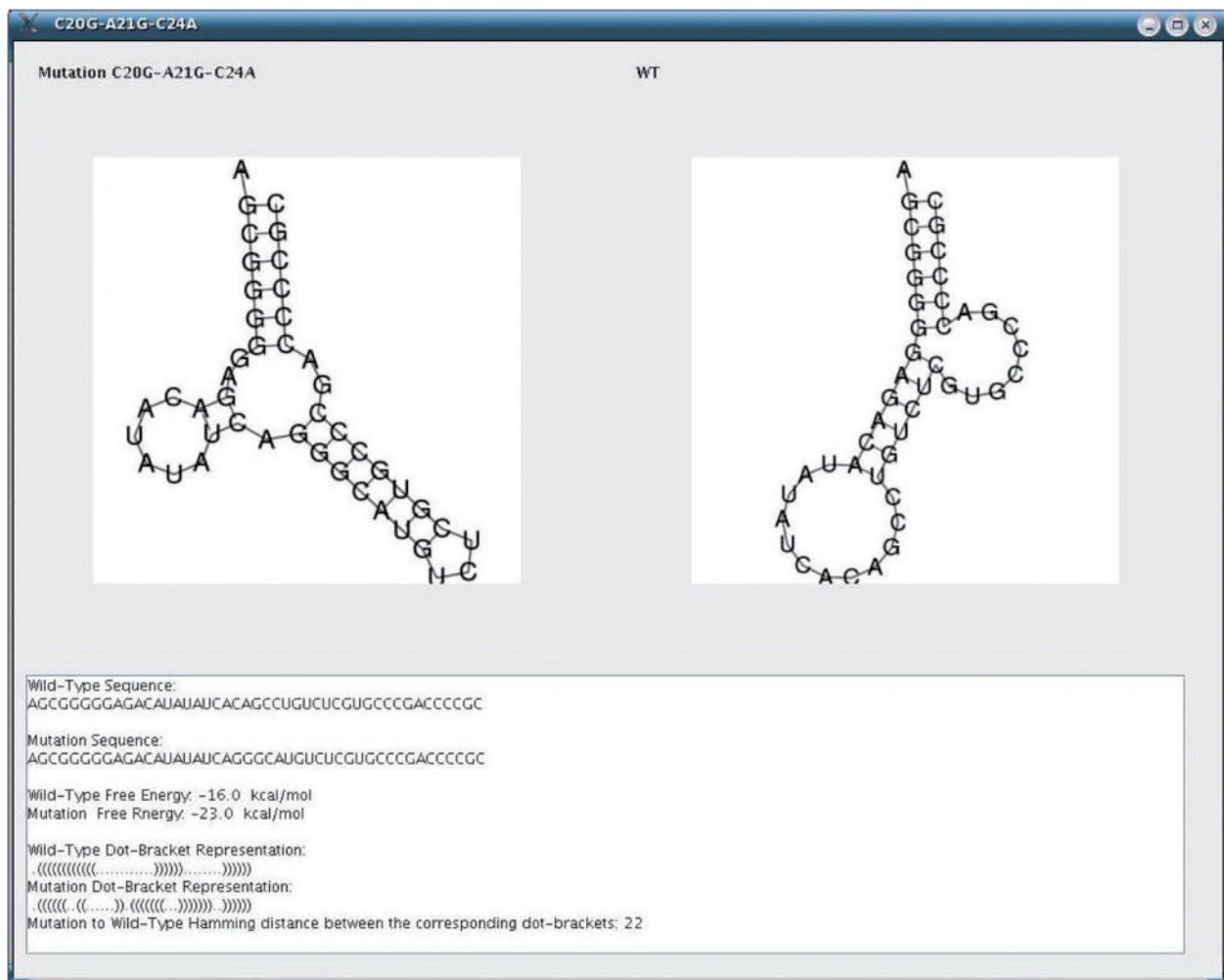


Figure 3: Output screen of a rearranging mutation in the 5BSL3.2 as a result of running RNAMute for the case of 3-point mutation and single clicking in the mutation group list screen shown in Figure 2 on the highlighted line corresponding to mutation C20 G-A21 G-C24A. The secondary structure drawings for the wild-type and the mutants are plotted.

mutations (choosing in the parameters: mutations = 2), there is an experimental confirmation of the prediction that is available in the gel of Figure 8 in [1].

RDMAS in detail

RDMAS [7] is a webserver for predicting single-point deleterious mutations on a given sequence and its predicted RNA secondary structure. Similar to RNAMute, it is also based in its core on programs from the Vienna RNA package [10, 22], such as RNAfold, RNAsubopt and RNAdistance. Given a sequence, it handles all single-point mutations in that sequence by a variety of similarity measures between the mutants and the wild-type predicted structures. There is a graphical output of the results, where the

maximum difference in structures between the wild-type and the possible mutants at each position are extracted into a structural deleteriousness profile that is plotted as a waveform. This way, the structurally important locations within the sequence can be easily detected by peaks with high structural deleteriousness on the profile.

First, the user can choose between an interactive job mode (for an RNA sequence <200 nt) and a batch job in which a correct email address is needed to be supplied. A job name can be specified or assigned automatically as explained in the RDMAS manual. Second, the user inserts an RNA sequence and chooses from a variety of methods to measure similarity between mutants and wild-type. A mode should be selected whether to consider suboptimal

solutions available with Vienna's RNAsubopt [23], and whether with maximum distance, minimum distance, or Boltzmann weighted distance as explained in the manual. Then a method is selected among four main choices and in each, sub-choices can be specified. The main choices are difference between free energy, edit distance, difference between topological indices, and base-pair distance. Sub-choices are varied and their description in more detail can be found in [7]. For example, among the topological indices, the second eigenvalue of the Laplacian matrix corresponding to a coarse-grained tree graph as suggested in [34] based on [35, 36] can be selected as a choice. At the same time, all standard distance measures including the commonly used base-pair distance are offered.

RNAmutants in detail

RNAmutants [8, 9] is a webserver for exploring the effects of mutations on the secondary structure of RNAs. It aims to provide a complete analysis of the mutational landscape for a given RNA sequence. In this context, a mutational landscape means that for a given sequence and a specified integer value k , the thermodynamic information that is being calculated by RNAmutants includes all the neighboring mutants of the given sequence up to k . RNAmutants is available both as a server and as a source code that can be downloaded on various Linux and Mac OS-X machines. In addition, Python code that can be useful for the creation of mutational profiles along with RNAmutants are available for download. In short, for a given sequence and a specified integer value k as mentioned above, a mutational profile is a 2D plot containing probabilities that allows us to identify fragile and robust positions in the sequence. The probabilities are mutation frequencies computed from sampled structures. More comprehensive details and some illustrations of mutational profiles are available in [8, 9].

The user first supplies a single RNA sequence along with an email address, an upper bound for the number of mutations allowed, and optionally the temperature in degrees Celsius. In addition, because the method employed by RNAmutants is based on a sampling procedure as described in [8], the desired number of sampled structures should also be supplied. Permitted values are in the range 0–1000 (default is 100). The GUI is friendly and depending on the sequence length and parameter values inserted, the user can wait for either an immediate answer or a calculation that may take several

hours and when finished the results are sent by email. The latter may occur if the sequence is long, or many mutations are specified in the upper limit, or many sampled structures are used, or a combination of these factors. Another example of an intensive calculation that also demands downloading RNAmutants is in the case of generating a mutational profile such as the ones mentioned above and illustrated in [8, 9].

The output from RNAmutants consists of several types of information. The first contains superoptimal solutions for the given sequence, where a superoptimal solution is that which has the lowest free energy over all secondary structures of all k -point mutants of the input RNA sequence. For example, if a value of 10 denotes the upper bound for the number of mutations allowed then for each 1-point, 2-point, ..., 10-point mutation, the value of the Boltzmann partition function is given (typically a large number), followed by the corresponding superoptimal sequence, structure, and mean free energy value in kcal/mol for each k -point mutation. Lower case letters in the sequence indicate nucleotide mutations in the input sequence, and Vienna's dot-bracket representation [22] is used to represent RNA secondary structures. The second contains sampled structures from Boltzmann ensemble, where at the top appears the initial sequence and below it are the mutated sequences and their corresponding secondary structures in dot-bracket representation. Finally, a graph showing the ensemble free energy in kcal/mol versus number of point-mutations (1-point, 2-point, ..., 10-point) is displayed, and additionally a mutational profile can be produced for the analysis of the complete mutation landscape but this may require an elaborate calculation and the usage of the Python programs available for download from the RNAmutants website.

For the same example sequence that was used above in describing RNAMute in detail, namely the HCV *cis*-acting replication element (5BSL3.2) taken from [1], the input sequence was inserted to the RNAmutants webserver along with an upper bound of 10 mutations and default parameters for temperature and sampling. After about 25 min, the superoptimal solutions appearing in Figure 4 were obtained, along with the sampled structures shown in Figure 5 (partial list). Downloading RNAmutants to a typical PC and running the same example took longer for the results to compute, approximately 40 min versus the 25 min reported above with the

```
# RNAmutants: part fun Z(k) and superoptimals MFE(k), 1
# 1
# AGCGGGGGGAGACAUAUAUCACAGCCUGUCUGUGCCCGACCCCGC
k      Z(k)      rnaSeq(k)      MFE(k)      Energy(k)
1      12091094768225.099609      AGCGGGGGGAGgCAUAUAUCACAGCCUGUCUGUGCCCGACCCCGC      .(((((((.((((((...
2      12374333986813600.000000      AGCGGGGGGAGACAgUAUAUCACAGCCUGUCUGUGCCCGACCCCGC      .((((((((((((((((...
3      3801703921289050112.000000      AGCGGGGGGAGgCAUgUAUgACAGCCUGUCUGUGCCCGACCCCGC      .(((((((.((((((...
4      951312923152816013312.000000      AGCGGGGGGAGgCAUgUAUgACAGCCUGUCUGUGCCCGcCCCGC      .(((((((.(((
5      175737490771789983252480.000000      AGCGGGGGgGgCAUgUAUgACAGCCUGUCUGUGCCCGcCCCGC      .(((((((((((
6      17076533579496199842430976.000000      AGCGGGGGgGgCAcgUAUgACAGCCUGUCUGUGCCCGcCCCGC      .(((((((((((
7      2253145781471559944098545664.000000      AGCGGGGGcGgCAUAUgagACAGCCUGUCUGUGgCCGcCCCGC      .((((
8      180290524045827010196599209984.000000      AGCGGGGGcGgGcUAUgagACAGCCUGUCUGUGCCCGcCCCGC      .((((
9      8053531710034269860325794775040.000000      AGCGGGGGcGgGcUcgAgCACAGCCUGUCUGAGCCCGcCCCGC      .((((
10     221183945285479982955280071655424.000000      AGCGGGGGcGgGcUcgAgCACAGCCUGgCUCGAGCCCGcCCCGC
```

Figure 4: Output screen from the RNAmutants webserver displaying superoptimal secondary structures, their free energy and mutation locations for the 5BSL3.2. Mutated nucleotides are shown in lower case. Each line contains the partition function value, the sampled mutated sequence, and the start of its minimum free energy structure in dot bracket notation concatenated for ease of display (missing are the continuation of the minimum free energy structure in dot bracket notation and the free energy of that structure, appearing in the right-most part of the RNAmutants output screen).

```
# Sampled structures from each k-Boltzmann ensemble for 1
# 1
# AGCGGGGGGAGACAUAUAUCACAGCCUGUCUGUGCCCGACCCCGC
AGCGGGGGGgGACcAcgAgCACAGCCUGUCUGUGgUCcCCCGC
.((((((((((((((((((((((((((((((((((((((((((((((((
AGCGGGGGGcGgCcUcgAUCACAGuCUgUgUCGaGgCCGcCCCGC
.((((((((((((((((((((((((((((((((((((((((((((((((
AGCGGGGGGgGgCAcgggCAGAGCCUGgCcCGUGCCCGACCCCGC
.(((((((.((((((((((((((((((((((((((((((((((((
AGCGGGGGGcGAGgCAUgCgACAGCCUGUCgCaUGCCCGcCCCGC
.((((((((((((((((((((((((((((((((((((((((((((((((
AGCGGGGGGgGuCAUAcgaCACAGCCUGUgUCUGGgACcCCCGC
.((((((((((((((((((((((((((((((((((((((((((((((((
AGCGGGGGGcGgCAUgagaCACgGagUGUCUGUGCCCGcCCCGC
.((((((((((((((((((((((((((((((((((((((((((((((((
AGCGGGGGGAGgCAcgccgCACAGCCUGcgCGUGCCCGcCCCGC
.(((((((.((((((((((((((((((((((((((((((((((((
cGCGGGGGGcGgCAcAgaCACAGCCUGUgUCUGGgCCGcCCCGC
.((((((((((((((((((((((((((((((((((((((((((((((((
AGCGGGGGGgGgCAUgCgCACAGCCUGcCgCGUGCCCGcCCCGC
.((((((((((((((((((((((((((((((((((((((((((((((((
gGCGGGGGGgGgCAUgaggCACAGCgUGUCUGUGCCCGcCCCGC
.((((((((((((((((((((((((((((((((((((((((((((((((
AGCGGGGGGcGgCcUcUgagACAGCCUGUCUGAGgCCGcCCCGC
.((((((((((((((((((((((((((((((((((((((((((((((((
AGCGGGGGGcGgCgCAUgUCgCAGCCUGcgCGUGCCCGcCCCGC
.((((((((((((((((((((((((((((((((((((((((((((((((
AGCGGGGGGgGAGgCAcAagACAGCCUGUCUgUGCCCGcCCCGC
.((((((((((((((((((((((((((((((((((((((((((((((((
```

Figure 5: Output screen from the RNAmutants webserver displaying initial portion of 100 mutant sequence/structure pairs for the 5BSL3.2. Mutated nucleotides are shown in lower case.

webserver. This demonstrates that the authors of RNAmutants have praiseworthy devoted considerable resources when putting up their server, which should be exploited by the users.

In [8], an analysis of this example is provided also with respect to its Rfam [37] consensus structure.

Figure 7 of [8] contains a list of pointwise mutants ordered by their break number, which is a measure of structural distortion that was devised in [8] for the purpose of identifying deleterious mutations. As possible improvements, neither the mutant structures in this list, nor any of the structures we observed in the

full list of sampled structures for which part of the list is available in Figure 5 contained the 2-point deleterious mutation found in Figures 18 and 19 of [6] that was reported in the gel provided in Figure 8 of [1]. This experimentally found 2-point deleterious mutation that its secondary structure can geometrically be represented as a star shaped coarse-grained tree-graph (containing two hairpins, see discussion in [6]) was detected as deleterious using RNAMute with the parameter value mutations = 2 in less than a minute, whereas no star shaped coarse-grained structures could be found when using 100 sampled structures with RNAmutants and only one such coarse-grained structure could be detected after increasing the number of sampled structures to 999, the maximum allowed. We assume that if we run RNAmutants with the maximum allowed sampled structures of 999 then eventually we may detect the aforementioned 2-point deleterious mutation after numerous multiple runs. However, such a procedure is not the standard way of utilizing RNAmutants that is expected from a user unaware of the experimental 2-point deleterious mutation in advance. Still, noteworthy of good mention is the analysis reported in [8] for the HCV *cis*-acting replication element with respect to its Rfam alignment. Such analyses may provide some interesting insights and may prompt some additional experiments to be tried.

Using the programs

The main task for using these programs is to insert an RNA sequence into one of them and analyze point mutations with relation to the wild-type sequence. All programs offer the possibility of additional parameters to be chosen by the user, with default values displayed at the beginning. As final output, all programs offer description of certain parts of the analysis as well. RNAMute directly displays the secondary structure drawings of the wild-type and mutant, which is easily accessible to the user by single-clicks from the mutation table. RDMAS displays a structural deleteriousness profile and deleteriousness distribution histogram. RNAmutants can display a mutation profile, after downloading and running the Python code as instructed in the manual. It is expected that more graphical outputs will be added as features in the future. In addition, the textual output provided in all these programs is substantially contributing to the analysis, as an essential step before the graphical output. The programs are user friendly

and can also be worked out by a non-specialist user along with the corresponding manuals and instructions that are available in the websites.

DISCUSSION

The programs RNAMute, RDMAS and RNAmutants have been developed in the past several years and offer some interesting prospects for mutational analysis of RNAs. They all in one way or the other, including RNAmutants, rely at present on thermodynamic parameters and therefore sequences that have been shown to fold correctly by experimental structure determination techniques to their energy minimization predicted structure are the best to work with as inputs to these programs in order to achieve reliable results. Though exceptional cases exist, in general the upper range estimate for the sequence length that these programs are useful for is ~ 150 nt, and for example RNA functional motifs of up to 150 nt that form stable stem-loop structures and are taken from UTRs or ORFs of RNA viruses may constitute favorable candidates for the use of these tools although this is by no means inclusive. It is expected that in future, the number of such RNA sequences will grow significantly and more biological systems involving RNAs will be analyzed by the aid of these programs.

Runtime can be a critical issue concerning the usage of these tools. Relying on an exhaustive listing of all mutation combinations may work fine for single-point mutations, but when dealing with multiple-point mutations the problem quickly become too heavy computationally. Thus, putting RDMAS aside because it is only capable of dealing with single-point mutations, a runtime comparison of the other two approaches is provided in Table 2. The times reported are in minutes. Standard parameter values were used in the comparison. The specific parameter values for each tool are available in the caption of Table 2. Although RNAMute can work in a faster mode if its method parameter is changed such that only stabilizing mutations of suboptimal solutions will be taken into account instead of stabilizing and destabilizing (of optimal solution) mutations, for fairness in the comparison the standard option was used while tuning the other parameter values according to the sequence length to achieve a fair timing comparison. Table 2 shows that while RNAMute works well for 1-point, 2-point and 3-point mutations, when k is further increased in

Table 2: Runtime comparison between RNAmutants and MultiRNAMute

	One mutation	Two mutations	Three mutations	Four mutations	Five mutations
RNAmutants					
50nt	0.03	0.8	0.28	1.02	2.87
100nt	0.45	1.09	3.64	14.7	78
150nt	0.75	3.5	12.33	96	284
MultiRNAMute					
50nt	0.03	0.12	0.92	6.33	54
100nt	0.08	0.32	2.5	18	216
150nt	0.4	1.62	32	>500	>500

The times are reported in minutes. For RNAmutants, a parameter value of 100 sampled structures is used. For MultiRNAMute, method = Fast, stabilizing and destabilizing is used, and in addition the following parameter values were chosen based on the input sequence length: for 50 nt, dist1 = dist2 = 35 and e = 7 were used; for 100 nt, dist1 = dist2 = 50 and e = 9 were used; for 150 nt, dist1 = dist2 = 75 and e = 9 were used; see text for explanation

the case of k -point mutations then RNAmutants has a better scaling behaviour. This is clearly shown in the case of the 150 nt sequence, where it is indicated that for the values of 4-point and 5-point mutations, the running time of RNAMute is above 500 min while RNAmutants becomes advantageous.

While RDMAS can handle single-point mutations in its webserver, both RNAMute and RNAmutants have been developed to analyze multiple point mutations [6, 8]. RNAMute does so by first applying Vienna's RNAsubopt [23], and then looking for mutations that stabilize the suboptimal solutions and destabilize the optimal one. Thus, RNAMute adheres to the way tracebacks are performed using RNAfold [10] and Mfold [11]. As in the example of the 5BSL3.2 of a subgenomic HCV replicon taken from [1] that was illustrated in the previous section, RNAMute succeeds to find the 2-point deleterious mutation reported in the gel shown in Figure 8 of [1], which was the result of the biochemical experiment. It also succeeds to find other deleterious mutations reported in the literature, and even to predict them before a biological experiment is performed to confirm the prediction [28]. Its strength is in analyzing, for example, 1-point, 2-point and 3-point mutations accurately for cases in which the wild-type secondary structure is well predicted by energy minimization. RNAmutants analyzes multiple-point mutations by a stochastic backtracking method similar to the one introduced in [38]. It is an interesting idea that offers good scalability (see Table 2) and allows large scale

simulations involving the calculation of all k -mutants, where k can be a large integer number, of a given RNA sequence. Such an approach, for example, may also be extended to automatically compute a neutral path in sequence space between two folds [39]. Already demonstrated applications include the detection of regions in an RNA sequence that have been optimized in the course of evolution [8]. These two approaches for the analysis of multiple-point mutations given an RNA sequence, namely the RNAMute approach [6] and the RNAmutants approach [8], are complementary to each other and can offer RNA biologists some needed computational tools. The demand for these programs is expected to increase as more research is advanced on RNA mutations.

Key Points

- RNA mutational analysis programs should be made user friendly and accessible to biologists as much as possible, both in terms of ease of use and simplification of the output such that it becomes understandable to the non-specialist.
- Experimental verifications whenever possible should be attempted in conjunction with using these programs.
- The various approaches for the analysis of multiple-point mutations should benefit from each other in advancing the field forward. Delivering their concept to biologists will help identify more systems for which their contribution will become rewarding.

FUNDING

Lynn and William Frankel Center for Computer Sciences at Ben-Gurion University, Israel; USA-Israel Binational Science Foundation (BSF grant 2003291).

References

1. You S, Stump DD, Branch AD, *et al*. 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**:1352–66.
2. 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**:2359–64.
3. Zimmerman JM, Maher LJ 3rd. In vivo selection of spectinomycin-binding RNAs. *Nucleic Acids Res* 2002;**30**:5425–35.
4. LeCuyer KA, Crothers DM. Kinetics of an RNA molecular switch. *Proc Natl Acad Sci USA* 1994;**91**:3373–7.

5. Churkin A, Barash D. RNAMute: RNA secondary structure mutation analysis tool. *BMC Bioinformatics* 2006;**7**:221.
6. Churkin A, Barash D. An efficient method for the prediction of deleterious multiple-point mutations in the secondary structure of RNAs using suboptimal folding solutions. *BMC Bioinformatics* 2008;**9**:222.
7. Shu W, Bo X, Liu R, *et al.* RDMAS: a web server for RNA deleterious mutation analysis. *BMC Bioinformatics* 2006;**7**:404.
8. Waldispühl J, Devadas S, Berger B, *et al.* Efficient algorithms for probing the RNA mutation landscape. *PLoS Comput Biol* 2008;**4**:e1000124.
9. Waldispühl J, Devadas S, Berger B, *et al.* RNAMutants: a web server to explore the mutational landscape of RNA secondary structures. *Nucleic Acids Res* 2009;**37**:W281–6.
10. Hofacker IL. Vienna RNA secondary structure server. *Nucleic Acids Res* 2003;**31**:3429–31.
11. Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003;**31**:3406–15.
12. Zuker M, Stiegler P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res* 1981;**9**:133–48.
13. Fox GE, Woese CR. 5S RNA secondary structure. *Nature* 1975;**256**:505–7.
14. Cannone JJ, Subramanian S, Schnare MN, *et al.* The Comparative RNA Web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. *BMC Bioinformatics* 2002;**3**:2.
15. Shapiro BA. An algorithm for comparing multiple RNA secondary structures. *Comput Appl Biosci* 1988;**4**:387–93.
16. Margalit H, Shapiro BA, Oppenheim AB, *et al.* Detection of common motifs in RNA secondary structures. *Nucleic Acids Res* 1989;**17**:4829–45.
17. Le SY, Nussinov R, Maizel JV. Tree graphs of RNA secondary structures and their comparisons. *Comput Biomed Res* 1989;**22**:461–73.
18. Shapiro BA, Zhang K. Comparing multiple RNA secondary structures using tree comparisons. *Comput Appl Biosci* 1990;**6**:309–18.
19. Nussinov R, Jacobson A. Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proc Natl Acad Sci* 1980;**77**(11):6309–13.
20. Nussinov R, Pieczenik G, Griggs JR, *et al.* Algorithm for loop matchings. *SIAM J Appl Math* 1978;**35**:68–82.
21. Zuker M. On finding all suboptimal folding of an RNA molecule. *Science* 1989;**244**:48–52.
22. Hofacker IL, Fontana W, Stadler PF, *et al.* Fast folding and comparison of RNA secondary structures. *Monatsh Chem* 1994;**125**:167–88.
23. Wuchty S, Fontana W, Hofacker IL, *et al.* Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers* 1999;**49**:145–65.
24. Mathews DH, Sabina J, Zuker M, *et al.* Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* 1999;**288**:911–40.
25. Barash D. Deleterious mutation prediction in the secondary structure of RNAs. *Nucleic Acid Res* 2003;**31**:6578–84.
26. Kass I, Horovitz A. Mapping pathways of allosteric communication in GroEL by analysis of correlated mutations. *Proteins: Struct Funct Genet* 2002;**48**:611–7.
27. Noivirt O, Eisenstein M, Horovitz A. Detection and reduction of evolutionary noise in correlated mutation analysis. *Protein Eng Des Sel* 2005;**18**:247–53.
28. Barash D, Gabdank I. Energy minimization methods applied to riboswitches: a perspective and challenges. *RNA Biol* 2010;**7**:89–96.
29. Churkin A, Barash D. Structural analysis of single-point mutations given an RNA sequence: a case study with RNAMute. *EURASIP J Appl Signal Processing* 2006;**56246**:1–7.
30. Churkin A, Elgad O, Petgho O, *et al.* RNAMute: RNA secondary structure mutation analysis tool. In: *Proceedings of the 4th International Computational Systems Bioinformatics Workshops* (CSB 2005) Stanford, CA: IEEE Computer Society Press, 2005. 139–41.
31. Clote P, Waldispühl J, Behzadi B, *et al.* Energy landscape of k-point mutants of an RNA molecule. *Bioinformatics* 2005;**21**:4140–7.
32. Höchsmann M, Voss B, Giegerich R. Pure multiple RNA secondary structure alignments: a progressive profile approach. *IEEE/ACM Trans Comput Biol Bioinform* 2004;**1**:53–62.
33. Steffen P, Voss B, Rehmsmeier M, *et al.* RNASHapes: an integrated RNA analysis package based on abstract shapes. *Bioinformatics* 2006;**22**:500–3.
34. Barash D. Second eigenvalue of the Laplacian matrix for predicting RNA conformational switch by mutation. *Bioinformatics* 2004;**20**:1861–69.
35. Fiedler M. Algebraic connectivity of graphs. *Czechoslovak Math J* 1973;**23**:298–305.
36. Merris R. Characteristic vertices of trees. *Lin Mult Alg* 1987;**22**:115–31.
37. Griffiths-Jones S, Bateman A, Marshall M, *et al.* Rfam: an RNA family database. *Nucleic Acids Res* 2003;**31**:439–41.
38. Ding Y, Lawrence CE. A Bayesian statistical algorithm for RNA secondary structure prediction. *Comput Chem* 1999;**23**:387–400.
39. Schultes EL, Bartel DP. One sequence, two ribozymes: implication for the emergence of new ribozyme folds. *Science* 2000;**289**:448–52.