

Structural Analysis of Single-Point Mutations Given an RNA Sequence: A Case Study with RNAMute

Alexander Churkin¹ and Danny Barash^{1,2}

¹Department of Computer Science, Ben-Gurion University, 84105 Beer-Sheva, Israel

²Genome Diversity Center, Institute of Evolution, University of Haifa, Israel

Received 2 May 2005; Revised 13 September 2005; Accepted 1 December 2005

We introduce here for the first time the RNAMute package, a pattern-recognition-based utility to perform mutational analysis and detect vulnerable spots within an RNA sequence that affect structure. Mutations in these spots may lead to a structural change that directly relates to a change in functionality. Previously, the concept was tried on RNA genetic control elements called “riboswitches” and other known RNA switches, without an organized utility that analyzes all single-point mutations and can be further expanded. The RNAMute package allows a comprehensive categorization, given an RNA sequence that has functional relevance, by exploring the patterns of all single-point mutants. For illustration, we apply the RNAMute package on an RNA transcript for which individual point mutations were shown experimentally to inactivate spectinomycin resistance in *Escherichia coli*. Functional analysis of mutations on this case study was performed experimentally by creating a library of point mutations using PCR and screening to locate those mutations. With the availability of RNAMute, preanalysis can be performed computationally before conducting an experiment.

Copyright © 2006 Hindawi Publishing Corporation. All rights reserved.

1. INTRODUCTION

The secondary structure of an RNA molecule is a representation of the pattern complementary base pairings that are formed between nucleic acids, given an initial RNA sequence. The sequence, represented as a string of four letters, is a single strand consisting of nucleotides A, C, G, U that folds according to minimum energy consideration as a basic assumption. The secondary structure of RNAs is experimentally accessible, thus making its computational prediction a challenging problem that can be tested in the laboratory. The folding prediction problem of the secondary structure of RNAs has been an area of active research since the late 70's (see [20] and other works, review available in [25]). Dynamic programming methods were developed in [15] (the Nussinov-Jacobson algorithm) for computing the maximum number of base pairings in an RNA sequence. Energy minimization methods by dynamic programming [23, 24] have led to Zuker's *mfold* prediction server [26] and the Vienna package [8]. An improvement in the success of these packages to predict an accurate folding comes from incorporating expanded energy rules [13], derived from an independent set of experiments, into the folding prediction algorithm. For sequences that are longer than approximately 150 nt, energy minimization methods may fail to reliably predict a secondary

structure from sequence alone. In those cases, an approach called comparative modeling [6] is preferable if it can be used.

In this paper, we address the problem of predicting desired nucleotide mutations, which relies on the success of RNA folding prediction by energy minimization but is independent of the particular folding algorithm itself. The question being asked is which nucleotide substitutions/deletions/insertions, introduced to the initial RNA sequence, will lead to a secondary structure rearrangement. The predictions are purely computational and can subsequently be tested in laboratory experiments. In order to validate our approach, we begin with an experimental result [22] that already succeeded to identify several selective mutations, inducing a conformation rearrangement in the secondary structure of RNA transcripts that inactivates spectinomycin resistance in bacteria. As a result, a concept that was initially proposed in [1] with analogy to computer vision scales is extended and applied for the inactivation of bacterial drug resistance. The method was previously tried to predict selective mutations in riboswitches and is here validated using results of an *in vivo* experiment performed independently.

Recently, much progress has been achieved towards understanding the function of small RNA structures in the control of important biological processes. From gene silencing

occurring in nature to nucleic acid engineering, in which innovative methods are being developed to modify or create new functional nucleic acids, the potential contribution of small RNAs to biotechnology and medical applications is evident. The possibility of causing drug resistance by the direct binding of short RNA transcripts with antibiotics, recently investigated in bacteria by *in vivo* selection experiments [22], is another advance in this field. We use this example discussed in [22] as our case study.

Selection experiments such as [22] demand adequate resources. A large pool of synthetic molecules with varying sequences needs to be created, before subjecting the pool to a desired selective pressure. Several repeated rounds of selection and amplification cycles are then applied. Oftentimes, without relation to a selection experiment, an interesting structure is obtained and its response to mutations leading to structural rearrangements can yield useful information on the properties of the structure itself. In such cases, because selection experiments are not performed on a regular basis as they demand planning and resources, computational prediction methods can help guide which mutations are worthwhile to explore further.

The paper is organized as follows. In Section 2, we introduce the notation and explain the motivation of using the Fiedler eigenvalue, or algebraic connectivity of trees, as a similarity measure between RNAfolds to locate structural rearrangements. We present some of the properties of the algebraic connectivity of trees that directly relate to the RNA mutation prediction problem. In Section 3, the general algorithm is presented for added layers of mutation (beyond single-point mutations). Section 4 provides numerical results for the prediction using the RNAMute package, followed by validation of the method using data from the laboratory experiment. Finally, Section 5 contains some concluding remarks and directions for further research.

2. RNA SIMILARITY WITH HIERARCHICAL STRUCTURES USING GRAPH SPECTRA

A similar concept that is used in computer vision to treat hierarchical structures (e.g., as reported recently in [16]) can be used to predict the effect of nucleotide mutations on the wildtype RNA secondary structure.

Let us examine the predicted secondary structure in Figure 1, as a result of running *mfold* [26] using dynamic programming to perform the energy minimization on pJ697 RNA [22], with the optimal solution shown in the figure. The folding prediction of the wildtype was used in [22] as a model to analyze the system behavior. The problem we are concerned with here is to predict the location of a mutation leading to conformational rearrangement. This can either be a single-point mutation, or if all single-point mutations are silent mutations, the least amount of consecutive nucleotide single-point mutations that will cause a structural transition. As a consequence of introducing the mutation, the new folded structure will assume a different shape from the wildtype secondary structure, signaling a structural transition that may disrupt or repair functional RNA motifs.

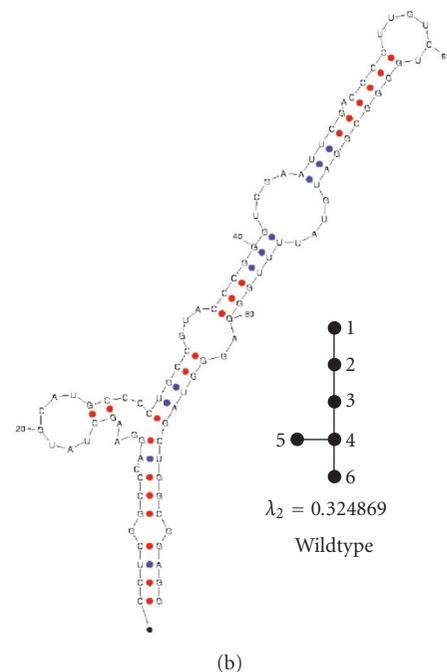
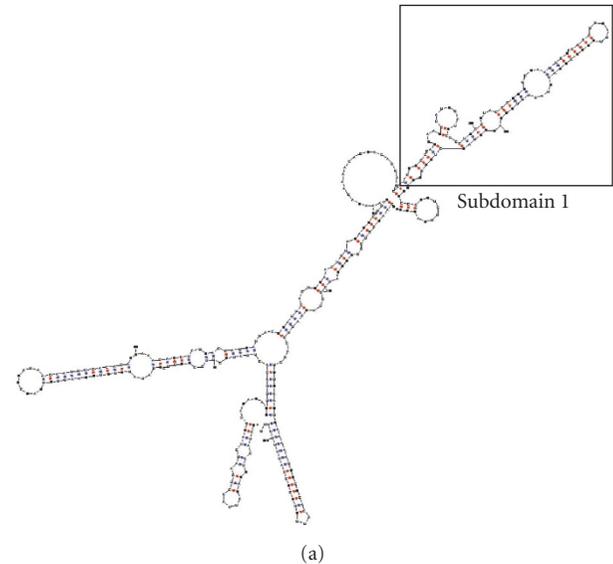


FIGURE 1: The predicted secondary structure of pJ697 RNA [22]. Subdomain 1 (boxed) is the region of interest for investigating conformational rearrangements that are thought to be responsible for the inactivation of spectinomycin resistance in *E. coli*. The predicted folding of subdomain 1 and its corresponding tree-graph representation, along with the Laplacian second eigenvalue, are also shown. Note that loops with single isolated nucleotides, by convention, are not accounted for as nodes in the tree-graph representation but the 5'-3' end is considered a node. Therefore we remain with exactly 6 vertices in the tree graph shown in Figure 1. Folding prediction of the boxed subdomain 1 by itself (right structure, labeled as wildtype) yields the same result as the folding prediction of the entire pJ697 RNA, extracting from it the secondary structure of subdomain 1.

For predicting selective mutations using the Laplacian second eigenvalue, as was suggested in [2], we use the algebraic connectivity of a tree as a similarity measure for comparing between the initial RNAfold and the folded structure of all possible mutants. The representation of RNA secondary structures as coarse-grained tree graphs was initially explored in [7, 11, 17] and the effect of single-point mutations using a combination of RNA tree-graph representation and string comparisons was addressed before in [12], without the reduction to eigenvalues with the methodology developed here. It should be noted that other similarity measures can be used (e.g., [9, 10, 18]) that convey more information about the RNA secondary structure representation by trees. The reduction into a coarse-grain tree-graph representation quantified by the algebraic connectivity of trees is simple and efficient. Moreover, it is easy to use the algebraic connectivity as a first-order approximation for the purpose of classification and filtering of unwanted structures when the information is arranged in a table, because of the favorable properties listed in the next section.

Let $T = (V, E)$ be a tree with vertex set $V = v_1, v_2, \dots, v_n$ and edge set E . Denote by $d(v)$ the degree of v , where $v \in V$ is a vertex of T . The Laplacian matrix of T (also known to be the difference of the diagonal matrix of vertex degrees $D(T)$ and the adjacency matrix $A(T)$ [3, 5]) is $L(T) = (a_{ij})$, where

$$a_{ij} = \begin{cases} d(v_i) & \text{if } i = j, \\ -1 & \text{if } v_i, v_j \in E, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

$L(T)$ is a symmetric, positive semidefinite, and singular matrix. The lowest eigenvalue of $L(T)$ is always zero, since all rows and columns sum up to zero. Denote by $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n = 0$ the eigenvalues of $L(T)$. The second smallest eigenvalue, λ_{n-1} , is called the algebraic connectivity [3] of T and labeled as $a(T)$. Some properties of $a(T)$ that are relevant to the application presented here will be mentioned below, following the calculation of $a(T)$ for the pj697 RNA secondary structure example depicted in Figure 1.

2.1. Laplacian representation of case study

The eigenvalues of the Laplacian matrix are independent of the chosen labeling for the nodes in the tree graph, which only amounts to interchanges of rows and columns. For a particular labeling of the tree-graph example in the boxed part (subdomain 1) of Figure 1, the corresponding Laplacian matrix $L(T)$ becomes

$$L = \begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & -1 & 3 & -1 & -1 \\ 0 & 0 & 0 & -1 & 1 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 \end{pmatrix}, \quad (2)$$

where $a(T)$ corresponding to the tree T of the wildtype structure in Figure 1 is 0.324869, in between a star of 6 vertices and a linear tree of 6 vertices.

The algebraic connectivity $a(T)$ possesses special properties that are advantageous for the RNA secondary structure mutation prediction application presented here.

Properties of algebraic connectivity for trees

Let $T = (V, E)$ be a tree on n vertices with algebraic connectivity $a(T)$. Then:

- (1) $0 \leq a(T) \leq 1$,
- (2) $a(T) = 0$ if and only if T is not connected,
- (3) $a(T) = 1$ if and only if $T = K_{1,n-1}$ is a star on n vertices (upper bound),
- (4) $a(T) = 2(1 - \cos(\pi/n))$ if and only if $T = P_n$ is a path (lower bound),

The algebraic connectivity $a(T)$, or the second eigenvalue of $L(T)$, is smallest but positive when the RNA secondary structure assumes a linear shape (a path) and becomes identically 1 when the RNA secondary structure assumes a star shape [3, 4, 14]. Although other possibilities exist to distinguish between tree topologies, the second eigenvalues of the coarse-grain tree graphs are nonexpensive to calculate for the small-sized matrices we are dealing with and possess intuitive meanings supported by mathematical theorems.

3. METHOD AND IMPLEMENTATION USING RNAMUTE

We use the algebraic connectivity $a(T)$ of a tree T to construct a stepwise procedure that attempts to locate the least number of mutations needed to disrupt an RNA motif, specifying their positions in the wildtype sequence as the final output. We note that simply visualizing the new structures obtained by performing the allowed mutations is not feasible in practice, unless we devise a procedure that enables us to inspect the structure of only selective mutants.

(1) Let N be the number of nucleotides in the given wildtype sequence. If $N > 100$, try subdividing the sequence into independently folded domains, such as subdomain 1 in Figure 1 (the folding prediction of this subdomain by itself is the same as the folding prediction of the whole sequence in that region). The subdivision, if necessary, is performed only once and is based on prior knowledge of the wildtype structure. Denote by N' the number of nucleotides in the artificial sequence, corresponding to the subdomain of interest.

(2) Serially or in parallel, run a folding prediction calculation (Zuker's *mfold* or Vienna RNAfold) for each of the $N' \times 3$ single-point mutants, since for each nucleotide there are 3 possible mutations. Extract the tree T corresponding to the secondary structure of each mutant in the form of a Laplacian matrix $L(T)$. Calculate the algebraic connectivity $a(T)$, which is the second eigenvalue of $L(T)$. Derive the number of vertices in T , how many mutants will assume the shape T (frequency of occurrence). Arrange the data in an eigenvalue table, as illustrated in Figure 2. Additional structure comparison measures and energy information can be added to the table in separate columns. The RNAMute package, which is currently under development and will be fully

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

Second eigenvalue	Number of vertices	wild type	Frequency
0.166717	8	-	2
0.186393	8	-	10
0.198062	7	-	9
0.225377	7	-	27
0.260323	7	-	21
0.267949	6	*	37
0.321720	7	-	2
0.324869	6	WT	140
0.381966	6	*	26
0.381966	5	-	4
0.381966	7	-	4
0.438447	6	*	1
0.518806	5	-	9

FIGURE 2: RNAMute screen output of one table categorization. Eigenvalue table for the prediction of single-point deleterious mutations in the subdomain (boxed) of pJ697 RNA [22]. The clustering to discrete eigenvalues enables to discriminate redundant folding possibilities and concentrate on predicting candidates for secondary structure conformation rearrangements that can cause inactivation of spectinomycin resistance in *E. coli*. An asterisk is marked whenever the same number of vertices as in the wildtype tree-graph structure occurs. Furthermore, not shown here, clustering to different ranges of coarse-grained tree-edit distances is performed in RNAMute, based on Shapiro and Zhang [18].

described elsewhere, also calculates other distance information such as Shapiro and Zhang's RNA tree distance [18] and the Vienna RNA distance [7].

(3) If all $N' \times 3$ single-point mutants correspond to the same tree T of the wildtype, add additional layers of mutation by extracting the tree T and calculating the features in Step (2) for each one of the $(N' \times 3)^2$ double-point mutations, then $(N' \times 3)^3$ triple-point mutations, ..., $(N' \times 3)^m$ m -point mutations, as necessary (see stopping criterion in next step).

(4) Repeat the previous step until $m = m^*$, where m^* is the minimal number of mutations needed so that at least one of the mutants folds to a tree which is different than T of the wildtype. Attempt to use prior information from step $i < j$ at step j , using data from the biology experiment if available, such that at step j only $(N' \times 3)^{m_j - m_i}$ folding calculations are needed instead of $(N' \times 3)^{m_j}$.

(5) When $m = m^*$, analyze the final eigenvalue table and in the case of RNAMute, interactively experiment with various eigenvalues that were calculated and stored. First, check the eigenvalues (i.e., visualize the predicted folded structure of mutants leading to this eigenvalue) that are furthest from the eigenvalue corresponding to the tree T of the wildtype. Second, check eigenvalues with different number of vertices than the wildtype, especially those with peculiarities (extreme number of vertices, low frequency of occurrence). When finding an interesting conformation rearrangement, go back from the artificial sequence with $N' < N$ nucleotides to the original sequence with N nucleotides and report the positions of the nucleotide mutations within the sequence, leading to that transition.

At the completion of these steps, we obtain predicted mutations that lead to conformation rearrangements and can

be tested in an experiment. The prescribed method is implemented using a computer package written in C and Java called RNAMute, which currently calculates all single-point mutations. In addition to eigenvalue information, RNAMute includes tables with distance measures available in the RNADistance module that is a part of the Vienna package [7, 8].

4. RESULTS OF CASE STUDY

We concentrate on predicting single-point mutations that will cause structural rearrangements with respect to the wildtype structure of RNA transcripts from pJ697 [22] depicted in Figure 1. The six single-point mutations in subdomain 1 of Figure 1, found by the selection experiment to inactivate spectinomycin resistance, are listed in Table 1. Another useful finding as a result of an in vitro experiment performed in [22] with radio-labeled transcripts corresponding to pJ697 and one of the inactivating point mutations (referenced as "mut 1") is the ability of a single-point mutation to alter the distribution of RNA conformers. This supports the hypothesis that a single-point mutation can lead to a secondary structure conformation rearrangement, which is responsible for a change in the function of the RNA. Therefore, if we predict possible mutations that are causing structural transitions in subdomain 1 of Figure 1, it is likely that those mutations are serious candidates to inactivate spectinomycin resistance in *E. coli*. One such mutation was experimentally found in [22].

We implemented Step (1) of the algorithm (previous section) by verifying that the folding prediction of subdomain 1 (Figure 1) is the same as the folding prediction of the whole sequence in that particular domain. Furthermore, we note that the six mutations reported in Figure 4 that alter the subdomain conformation also alter the full RNA conformation as verified using *mfold*. Thus, our assumption that the subdomain of Figure 1 is an independently folded domain is likely to hold in the case study examined here. Consequently, our artificial structure for the purpose of mutation prediction consists only of the boxed segment in Figure 1 which is 97 nt long. Performing Step (2), the RNAMute package automatically generates an eigenvalue table for all $97 \times 3 = 291$ single-point mutations, depicted in Figure 2. In this case, since there is a large amount of single-point mutations leading to structural rearrangements, we stop the procedure described in the previous section at $m^* = 1$.

4.1. Analysis with RNAMute

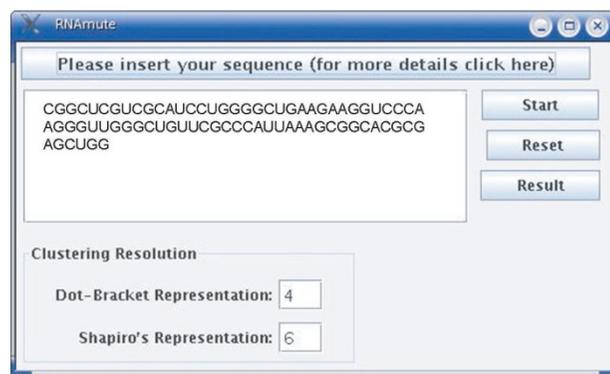
Figure 2 lists the structural rearrangement predictions of all possible single-point mutations, ranked by their second eigenvalue of the Laplacian matrix corresponding to the tree-graph representation of their folding prediction. It is expected that some of these folded structures will not occur in nature. We would like to examine how many of the inactivating mutations found by the experiment (Table 1) match various eigenvalues listed in Figure 2 and whether, provided we only have the information in Figure 2, we could have suggested meaningful mutations to test as candidates for

TABLE 1: Six single-point mutations in the subdomain (boxed) of pJ697 RNA [22] that inactivate spectinomycin resistance in *E. coli*, obtained by a selection experiment. From the observations in [22] it is likely that a conformation rearrangement in the secondary structure is associated with the inactivation. WT stands for the wildtype, the six nucleotide mutations are highlighted with the shaded boxes.

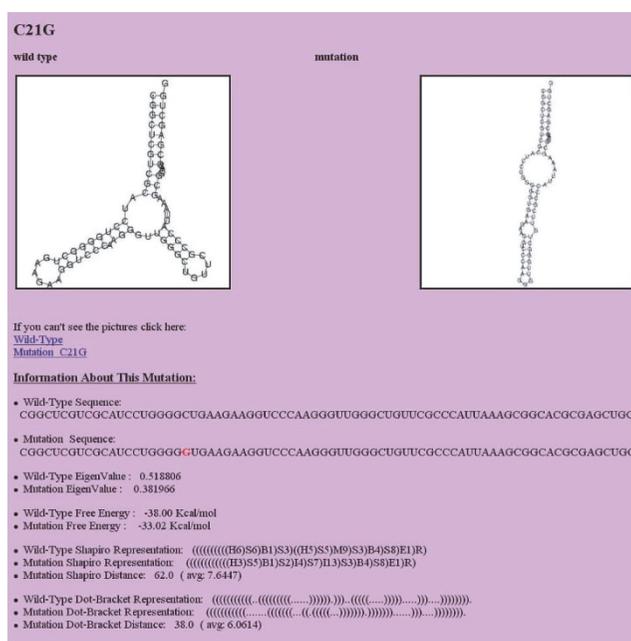
Mutation	Sequence
WT	CCUCGGCCCAGGAAGCUAUGCAUGC CCCUGCCGUACCCGGGUCGAAUUCG ACCCCUUGUCUGGGGCGGAUGUAUU UUGGGAGGGUAGCUGGCGGAGG
1	CCUCGGCCCAGGAAGCUAUGCAUGC CCCUGCCGUACCCGGGUCGAAUUCG ACCCCUUGUC C GGGGCGGAUGUAUU UUGGGAGGGUAGCUGGCGGAGG
2	CCUCGGCCCAGGAAGCUAUGCAUGC CCCUGCCGUACCCGGGUCGAAUUCG ACCCCUUGUCUGG A GCGGAUGUAUU UUGGGAGGGUAGCUGGCGGAGG
3	CCUCGGCCCAGGAAGCUAUGCAUGC CCCUGCCGUACCCGGGUCGAAUUCG ACCCCUUGUCUGGGGCGGAUGUAUU U A GGGAGGGUAGCUGGCGGAGG
4	CCUCGGCCCAGGAAGCUAUGCAUGC CCCUGCCGUACCCGGGUCGAAUUCG ACCCCUUGUCUGGGGCGGAUGUAUU UUGGGAGGG A AGCUGGCGGAGG
5	CCUCGGCCCAGGAAGCUAUGCAUGC CCCUGCCGUACCCGGGUCGAAUUCG ACCCCUUGUCUGGGGCGGAUGUAUU UUGGGAGGGU G GCUGGCGGAGG
6	CCUCGGCCCAGGAAGCUAUGCAUGC CCCUGCCGUACCCGGGUCGAAUUC A ACCCCUUGUCUGGGGCGGAUGUAUU UUGGGAGGGUAGCUGGCGGAGG

inactivating mutations in an experiment. Selection experiments are biased and thus they are likely to miss interesting mutations that can potentially be predicted using computer simulations.

For each of the six inactivating mutations in Figure 4, we simulate a folding prediction using *mfold*/Vienna (as was performed for “mut 1” in [22]). We then calculate the eigenvalue associated with that folding. Figure 4 captures the five distinct tree graphs corresponding to the six inactivating mutations and their associated eigenvalues. Examining Figure 2, it is noted that although the wildtype structure and mutations 1, 2, 5 fall into the same eigenvalue, their overall structure is different. For example, while mutations 1, 2 possess a multibranch loop and two hairpins, the wildtype possesses a single hairpin, although their tree graph compactness (hence second eigenvalue) is the same. To relieve this ambiguity, we further subdivide the tree-graphs associated with the same second eigenvalue into various groups according to their edit distances as suggested in Shapiro and Zhang [18]



(a)



(b)

FIGURE 3: RNAMute screen output of one single-point mutation, U77A of the full sequence, used in our case study example. Information includes the minimal energies of the wildtype and mutant, their sequences, their secondary structure representation in the Vienna dot-bracket notation and Shapiro’s coarse-grain string notation, and the distances between the two structures using Vienna’s RNAdistance and Shapiro’s tree-edit distance.

and available in our RNAMute implementation. Class (A) are mutations possessing “Shapiro distances” [7] in the range of 0–20 with respect to the wildtype, corresponding to a tree graph that is considerably close to the wildtype structure with respect to edit operations. Class (B) are mutations possessing “Shapiro distances” in the range of 81–99 with respect to the wildtype, corresponding to a tree graph surrounding mutations 1, 2. Class (C) are mutations possessing “Shapiro distances” in the range of 21–56 with respect to the wildtype,

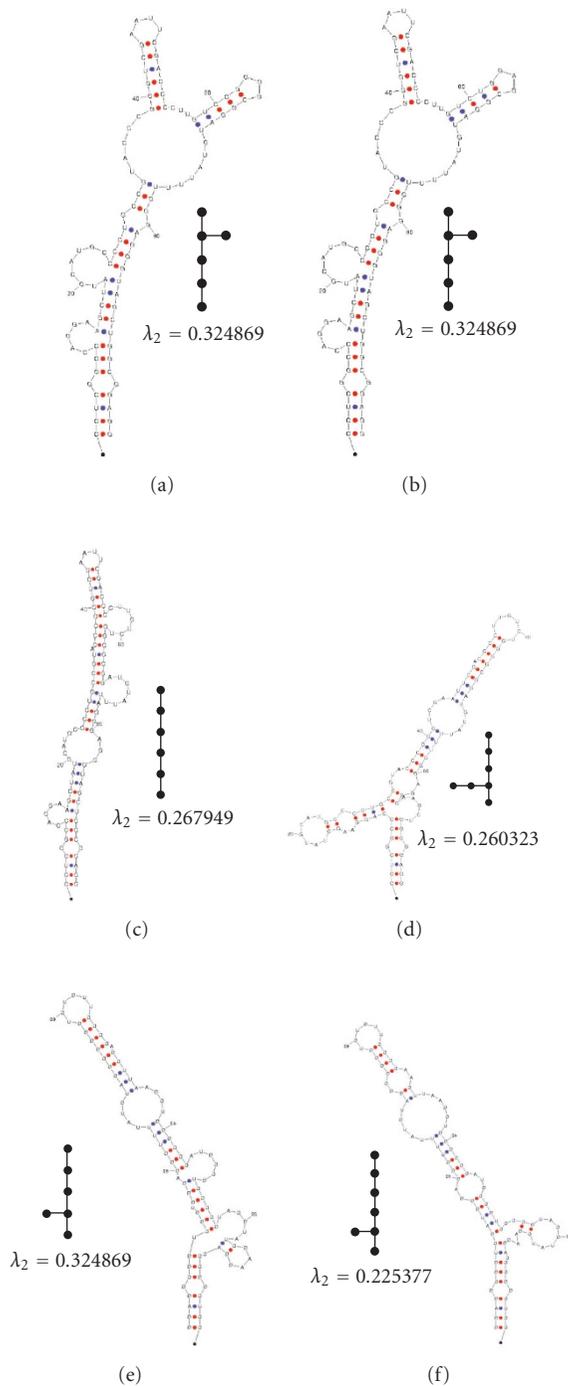


FIGURE 4: The secondary structure of the six mutants from Table 1, found in [22] to inactivate spectinomycin resistance in *E. coli* by a selection experiment. Their tree-graph representation and associated eigenvalues are drawn.

corresponding to a tree graph surrounding mutation 5. Thus, our analysis includes various measures to estimate similarity of secondary structures, a strategy that is taken in RNAMute. Furthermore, from Figure 2 we observe possibilities for peculiar mutant structures, such as a linear-shaped tree

graph with 8 vertices corresponding to $\lambda_2 = 0.166717$. Its low frequency of occurrences (two mutations out of any possible single-point mutations) is not necessarily an indication for false positives; a selection experiment may have skipped these mutations that are highly interesting to try in additional experiments. Such mutations are candidates for vulnerable spots in the wildtype sequence, potentially triggering a conformational switch that will lead to even stronger inactivation of spectinomycin resistance. Thus, our analysis with RNAMute (see Figure 3) can detect patterns that are worth exploring in additional laboratory experiments.

4.2. Other case studies

The case study reported in this paper [22] was the first we analyzed with RNAMute. Based on the gathered results, we have tried other test cases that require less assumptions to be made prior to predictions. A class of such test cases that will be reported in the future can potentially be used for the examination of phenotypic data available from hepatitis C virus (HCV) experiments [19, 21]. For example, RNAMute was able to single out a conformational rearranging mutation in the 5BSL3.2 structure that was reported experimentally in [21]. These test cases are shorter in their sequence lengths (< 100 nt), and they can be analyzed independently without further assumptions.

5. CONCLUSIONS

We have presented a method and its RNAMute package implementation for predicting nucleotide mutations that may intervene with RNA function through conformation rearrangements in the secondary structure. Admittedly, the method has several limitations, such as relying on the accuracy of energy minimization methods and the use of a coarse-grained measure. For longer sequences, this approach may fail, unless there are associated cases in which comparative modeling [6] can be used. Still, for some sequences it has already been shown to match experimental results (e.g., the *leptomonas collosoma* mentioned in [2]) and our recent RNAMute implementation includes fine-grain measures as well. The method is demonstrated on a case study by matching the prediction results with known point mutations that inactivate spectinomycin resistance in bacteria, obtained by a selection experiment [22]. Comparison of predicted mutations with the ones found by the experiment demonstrates the potential of the method. Thus, it can be used on a variety of RNA structures before planning an in vivo experiment, to detect vulnerable spots and suggest mutations that are interesting for further exploration.

ACKNOWLEDGMENTS

We thank James Maher from Mayo Clinic for his valuable comments and feedback to our work. The research was supported by a Grant from the Israel-USA Binational Science Foundation (BSF) 2003291.

REFERENCES

- [1] D. Barash and D. Comaniciu, "A common viewpoint on broad kernel filtering and nonlinear diffusion," in *Proceedings of the 4th International Conference on Scale-Space Theories in Computer Vision (Scale-Space '03)*, vol. 2695 of *Lecture Notes in Computer Science*, pp. 683–698, Isle of Skye, UK, June 2003.
- [2] D. Barash, "Second eigenvalue of the Laplacian matrix for predicting RNA conformational switch by mutation," *Bioinformatics*, vol. 20, no. 12, pp. 1861–1869, 2004.
- [3] M. Fiedler, "Algebraic connectivity of graphs," *Czechoslovak Mathematical Journal*, vol. 23, pp. 298–305, 1973.
- [4] R. Grone and R. Merris, "Algebraic connectivity of trees," *Czechoslovak Mathematical Journal*, vol. 37, no. 4, pp. 660–670, 1987.
- [5] R. Grone, R. Merris, and V. S. Sunder, "The Laplacian spectrum of a graph," *SIAM Journal on Matrix Analysis and Applications*, vol. 11, no. 2, pp. 218–238, 1990.
- [6] R. R. Gutell, J. C. Lee, and J. J. Cannone, "The accuracy of ribosomal RNA comparative structure models," *Current Opinion in Structural Biology*, vol. 12, no. 3, pp. 301–310, 2002.
- [7] I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, and P. Schuster, "Fast folding and comparison of RNA secondary structures," *Monatshefte für Chemie*, vol. 125, no. 2, pp. 167–188, 1994.
- [8] I. L. Hofacker, "Vienna RNA secondary structure server," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3429–3431, 2003.
- [9] T. Jiang, G. Lin, B. Ma, and K. Zhang, "A general edit distance between RNA structures," *Journal of Computational Biology*, vol. 9, no. 2, pp. 371–388, 2002.
- [10] J. Kitagawa, Y. Futamura, and K. Yamamoto, "Analysis of the conformational energy landscape of human snRNA with a metric based on tree representation of RNA structures," *Nucleic Acids Research*, vol. 31, no. 7, pp. 2006–2013, 2003.
- [11] S.-Y. Le, R. Nussinov, and J. V. Maizel, "Tree graphs of RNA secondary structures and their comparisons," *Computers and Biomedical Research*, vol. 22, no. 5, pp. 461–473, 1989.
- [12] H. Margalit, B. A. Shapiro, A. B. Oppenheim, and J. V. Maizel, "Detection of common motifs in RNA secondary structures," *Nucleic Acids Research*, vol. 17, no. 12, pp. 4829–4845, 1989.
- [13] D. H. Mathews, J. Sabina, M. Zuker, and D. H. Turner, "Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure," *Journal of Molecular Biology*, vol. 288, no. 5, pp. 911–940, 1999.
- [14] R. Merris, "Characteristic vertices of trees," *Linear and Multilinear Algebra*, vol. 22, pp. 115–131, 1987.
- [15] R. Nussinov and A. B. Jacobson, "Fast algorithm for predicting the secondary structure of single-stranded RNA," *Proceedings of the National Academy of Sciences*, vol. 77, no. 11, pp. 6309–6313, 1980.
- [16] A. Shokoufandeh, D. Macrini, S. Dickinson, K. Siddiqi, and S. W. Zucker, "Indexing hierarchical structures using graph spectra," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 7, pp. 1125–1140, 2005, Special issue on syntactic and structural pattern recognition.
- [17] B. A. Shapiro, "An algorithm for comparing multiple RNA secondary structures," *Computer Applications in the Biosciences*, vol. 4, no. 3, pp. 387–393, 1988.
- [18] B. A. Shapiro and K. Zhang, "Comparing multiple RNA secondary structures using tree comparisons," *Computer Applications in the Biosciences*, vol. 6, no. 4, pp. 309–318, 1990.
- [19] D. B. Smith and P. Simmonds, "Characteristics of nucleotide substitution in the hepatitis C virus genome: constraints on sequence change in coding regions at both ends of the genome," *Journal of Molecular Evolution*, vol. 45, no. 3, pp. 238–246, 1997.
- [20] M. S. Waterman and T. F. Smith, "RNA secondary structure: a complete mathematical analysis," *Mathematical Biosciences*, vol. 42, no. 3-4, pp. 257–266, 1978.
- [21] S. You, D. D. Stump, A. D. Branch, and C. M. Rice, "A cis-acting replication element in the sequence encoding the NS5B RNA-dependent RNA polymerase is required for Hepatitis C virus RNA replication," *Journal of Virology*, vol. 78, no. 3, pp. 1352–1366, 2004.
- [22] J. M. Zimmerman and L. J. Maher III, "In vivo selection of spectinomycin-binding RNAs," *Nucleic Acids Research*, vol. 30, no. 24, pp. 5425–5435, 2002.
- [23] M. Zuker and P. Stiegler, "Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information," *Nucleic Acids Research*, vol. 9, no. 1, pp. 133–148, 1981.
- [24] M. Zuker and D. Sankoff, "RNA secondary structures and their prediction," *Bulletin of Mathematical Biology*, vol. 46, no. 4, pp. 591–621, 1984.
- [25] M. Zuker, "Calculating nucleic acid secondary structure," *Current Opinion in Structural Biology*, vol. 10, no. 3, pp. 303–310, 2000.
- [26] M. Zuker, "Mfold web server for nucleic acid folding and hybridization prediction," *Nucleic Acids Research*, vol. 31, no. 13, pp. 3406–3415, 2003.

Alexander Churkin received his B.S. degree with distinction from the Department of Computer Science at Ben-Gurion University in 2004. Since September 2004, he has been a graduate student in the Department of Computer Science at Ben-Gurion University. His research interests include bioinformatics, RNA structure predictions, and scientific computing.



Danny Barash received his Ph.D. degree in applied science in 1999 from the University of California at Davis. From 1999 to 2001, he was employed at Hewlett Packard Laboratories in the Technion, Israel, pursuing research on image processing and computer vision. From 2001 to 2003, he was a Howard Hughes Medical Institute Postdoctoral Fellow at New York University and a Research Fellow at the Institute of Evolution in the University of Haifa, Israel, where he made a transition to computational biology. Since 2004, he has been with the Department of Computer Science at Ben-Gurion University, where he is currently an Assistant Professor in bioinformatics. His secondary affiliation is with the Institute of Evolution at Haifa University. His research interests include computational biology, RNA structure predictions, computational imaging, and numerical analysis.

