

Energy minimization methods applied to riboswitches

A perspective and challenges

Danny Barash* and Idan Gabdank

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

Key words: energy minimization methods, RNA folding prediction, riboswitches

Abbreviations: TPP, thiamin pyrophosphate

Energy minimization methods for RNA secondary structure prediction have been used extensively for studying a variety of biological systems. Here, we demonstrate their applicability in riboswitch studies, exemplified in both the expression platform and aptamer domains. In the expression platform domain, energy minimization methods can be used to predict *in silico* a unique point mutation positioned in the non-conserved region of the TPP riboswitch that will transform it from a termination to an anti-termination state, thus backing the prediction experimentally. Furthermore, a successive prediction can be made for a compensatory mutation that is positioned over half the sequence length of the riboswitch from the original mutation and that completely overturns the anti-termination effect of the original mutation. This approach can be used to computationally predict rational modifications in riboswitches for both research and practical applications. In the aptamer domain, energy minimization methods can be used when attempting to detect a novel purine riboswitch in eukaryotes based on the consensus sequence and structure of the bacterial guanine binding aptamer. In the process, some interesting candidates are identified, and although they are attractive enough to be tested experimentally, they are not detectable by sequence based methods alone. These brief examples represent the important lessons to be learned as to the strengths and limitations of energy minimization methods. In light of our growing knowledge in the energy minimization field, future challenges can be advanced for the rational design of known riboswitches and the detection of novel riboswitches. Unlike analyses of specific cases, it is stressed that all the results described here are predictive in scope with direct applicability and an attempt to validate the predictions experimentally.

Introduction

The discovery of riboswitches, recognized as a breakthrough in the understanding of RNA regulation, has opened the door to the development of bioinformatics methods to search for and

analyze the riboswitches. Although the discovery itself was biological, with the first experimental validations published in 2002 independently by the Breaker group and the Nudler group,¹⁻⁴ conserved sequence patterns in the 5' UTRs of bacteria were identified before by comparative analysis of the upstream regions of several genes expected to be co-regulated. These works contributed to the description of the RFN element,⁵ the S-box⁶ and the THI-box.⁷

Although the perspective given here is specifically about the applicability of energy minimization methods for riboswitch studies, mainly by the use of mfold,⁸ UNAFold,⁹ and the Vienna RNA package,^{10,11} some bioinformatics contributions since 2002 should be noted, although the list is by no means inclusive. A simple search program called Sequence Sniffer was used to discover a eukaryotic riboswitch.¹² Additional works on a genomic scale by Barrick et al.^{13,14} triggered many more findings of bacterial riboswitches. Ruzzo and coworkers, through their covariance models (CM) approach and the development of CMfinder,^{15,16} helped discover new classes of riboswitches.¹⁷⁻¹⁹ In addition, the insertion of known riboswitches into the RFAM database^{20,21} was instrumental in advancing the field. Gelfand and coworkers²² have continued advancing the comparative analysis approach in metagenomes. A simpler sequence based method²³ and a more sophisticated sequence based bioinformatics search method for detecting new riboswitches was developed in.²⁴ Another approach is that of,²⁵ which by stochastic context-free grammar, can also be used to search for new riboswitches. On the topic of searching for new riboswitches in genomes, a review is available in.²⁶ Regarding energy minimization methods, in addition to their application as an aid in the general description of the riboswitch structure used in,^{2,27} some prediction and analysis works that were assisted by these methods include.²⁸⁻³⁸

Results

Before examining specific applications of energy minimization methods to the design of and search for riboswitches, it is worthwhile noting that these methods are able to capture the on/off states (e.g., termination and anti-termination) by inserting the riboswitch sequence. This is not a trivial assertion in complex RNAs that are >200 nt long, such as, for example, the TPP

*Correspondence to: Danny Barash; Email: dbarash@cs.bgu.ac.il
Submitted: 10/08/09; Revised: 11/15/09; Accepted: 11/16/09
Previously published online:
www.landesbioscience.com/journals/mabiology/article/10657

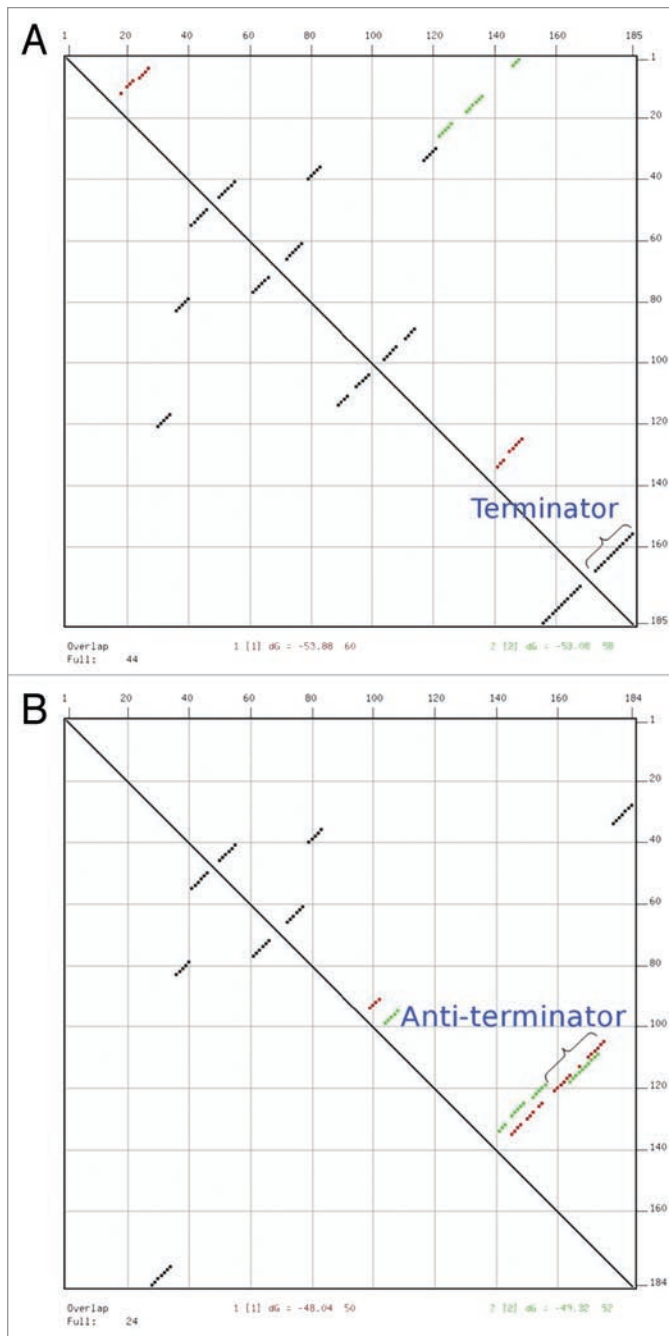


Figure 1. Energy dot plots of the TPP riboswitch. Mfold⁸ was used for illustration, and a similar result can be obtained using RNAfold from the Vienna RNA package.¹¹ (A) The dot plot showing the highly stable terminator stem. Although the anti-termination state is present, it is covered by the termination state. (B) The dot plot showing the antiterminator stem, after the terminator stem was removed by forcing the constraint P 161–168 173–180 in mfold.

riboswitch. There are limitations when predicting the RNA secondary structure of RNA sequences that are above 100–150 nt using energy minimization methods, with energy rules that are experimentally derived from much smaller combinations. For specific RNAs such as RNA viruses, one can still hope to retain relative accuracy beyond these nt limitations because of the

simplified nature of the elongated secondary structure. However, riboswitches have complex RNA secondary structures and do not possess such favorable properties.

Nevertheless, it turns out that even though not every base pairing in the secondary structure of a riboswitch >200 nt long can be predicted correctly, the two distinct on/off states of a riboswitch (e.g., termination and anti-termination) can be captured by energy minimization methods, a property that was first demonstrated in² using secondary structure drawings. More advanced techniques are described here. Clote and coworkers have devised a method called RNAbor^{31,32} whereby inspections of the Boltzmann probability density plot enabled them to detect possible conformational switches as in riboswitches. Here, we illustrate this possibility more simply for a particular case by using mfold⁸ and examining the energy dot plot, noting that the same procedure can be done using the Vienna RNA websuite³⁹ with a probability dot plot. Although RNAbor is a more sophisticated and specialized method for this task, our aim is only to demonstrate the concept that energy minimization methods are capable of detecting a riboswitch in a particular case. Therefore, the TPP riboswitch of² is used for illustration. When its sequence GCA GAA CAA UUC AAU AUG UAU UCG UUU AAC CAC UAG GGG UGU CCU UCA UAA GGG CUG AGA UAA AAG UGU GAC UUU UAG ACC CUC AUA ACU UGA ACA GGU UCA GAC CUG CGU AGG GAA GUG GAG CGG UAU UUG UGU UAU UUU ACU AUG CCA AUU CCA AAC CAC UUU UCC UUG CGG GAA AGU GGU UUU UUU is inserted into mfold and an energy dot plot is generated, the terminator stem is clearly visible (Fig. 1A). Using constraint information, it is possible to prohibit 8 base-pairings in the terminator stem by inserting the constraint ‘P 161–168 173–180’ and performing another energy minimization run, after which the antiterminator stem clearly appears (Fig. 1B). In principle, every RNA sequence can be examined as such, to the first order, to test whether it possesses a conformational switch. For automatic quantification and detection techniques of RNA sequences that possess this property, approaches like RNAbor^{31,32} can be considered.

Predicting unique point mutations in the expression platform domain. Correctly predicting the effect of unique point mutations on the structure and function of RNA is of primary interest to basic and applied research. Computational methods have been developed, therefore, to assist in designing such mutations.^{28,40} By examining the TPP riboswitch of² we focus on point mutations that can potentially affect transcription termination in bacteria. In addition to attempting to computationally predict riboswitch variants with desirable properties, aiming to assist in the design of experimental work for engineering artificial riboswitches^{41,42} (see Discussion), it is also important from the bioinformatics and long-range research standpoints, however, to develop a general tool that can reliably predict point mutations that cause conformational rearrangements^{43,44} (other tools for similar purposes are described in^{45–47}). Such work may also shed light on the properties of an RNA allosteric system.

The TPP riboswitch fits the criteria of such a model system that should be examined. It works at both the transcription

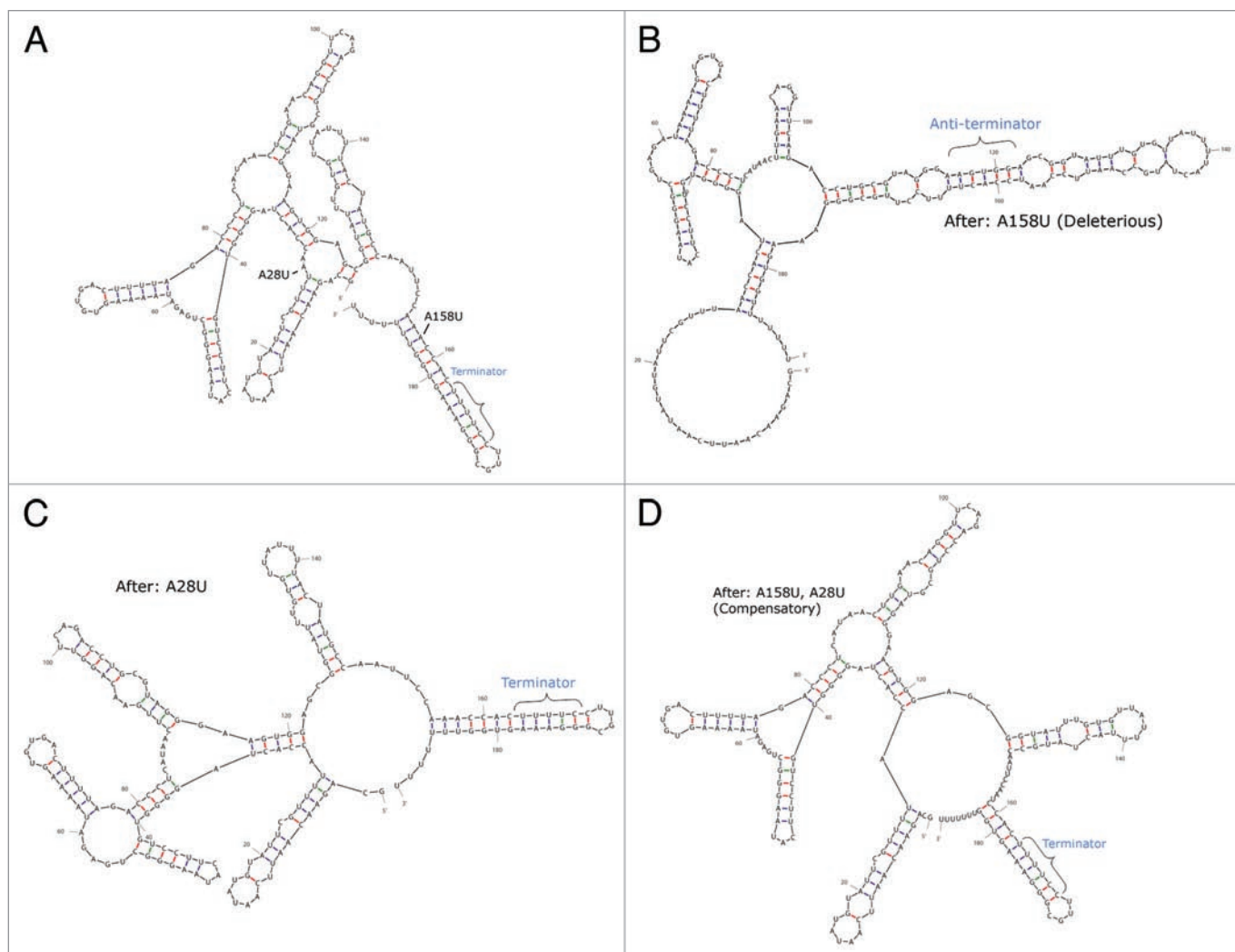


Figure 2. Thiamin pyrophosphate (TPP) riboswitch in transcription termination of *Bacillus subtilis*: wild type and mutants. (A) The wild type secondary structure according to the riboswitch model² possesses a stable terminator hairpin downstream from the evolutionary conserved bases in the thi-box. The anti-antiterminator structure that prohibits the emergence of an antiterminator structure as in (B), guarding the terminator hairpin in the presence of TPP, is located upstream to the terminator. (B) Deleterious mutated riboswitch. Predicted mutation A158U, with respect to the wild type structure in (A), leads to the emergence of the same antiterminator structure as in the absence of TPP. (C) Examination of predicted mutation A28U by itself, showing no effect on the terminator structure. (D) Compensatory mutated riboswitch. Predicted mutation A28U, with respect to the deleteriously mutated riboswitch in (B), restores the terminator hairpin and anti-antiterminator structures to yield a secondary structure identical to that of the wild type. Note the large distance between positions 158 and 28, making the prediction A158U, A28U non-obvious to obtain and of particular interest to research. UNAFold⁹ version 3.6 with mfold utils version 4.5 was used to predict and generate the secondary structure drawings.

termination² and translation initiation³ levels. The riboswitch comprises a conserved thi-box region⁴⁸ that binds a small molecule (TPP) and a non-conserved region that undergoes a conformational change to interrupt transcription or inhibit translation of thiamin biosynthetic genes. The riboswitch mechanism has been described in detail at the level of transcription termination.² The switch between termination and anti-termination states can be illustrated by a structural model² that uses Zuker's mfold⁸ or the Vienna package,¹¹ both of which rely on available energy parameters.⁴⁹ This simplified model forms the basis for the computational prediction of deleterious mutations.²⁸ Here, the prediction of deleterious (anti-termination) mutation A158U is supplemented by a compensatory mutation A28U, and both are verified experimentally. A large

(>120 nt) distance separates the two mutations, thereby making their computational prediction non-trivial. The locations of these mutations are depicted in Figure 2, which was generated with UNAFold,⁹ and the experimental results used to validate them are described in Materials and Methods and shown in Figure 3. The results are discussed in the following section.

Predicting bacterial consensus structures in the aptamer domain. Using energy minimization methods, a moving window approach can be used to scan regions of interest in the genomes of a variety of organisms. Each secondary structure prediction in the window can then be compared to the bacterial consensus structure of the aptamer domain. This method was employed in³⁷ where it is described in detail.

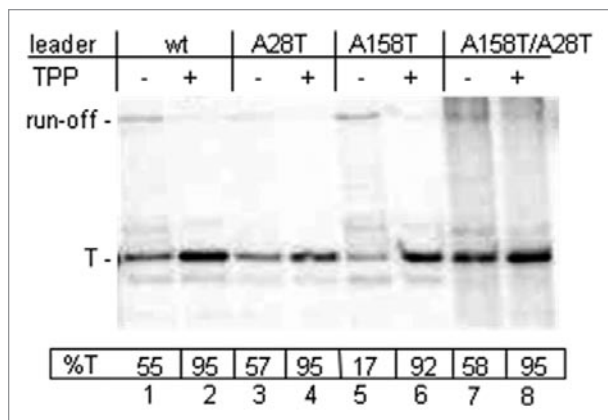


Figure 3. Effect of thi-leader mutations on termination. The autoradiogram of the 12% sequencing PAGE shows [³²P]-labeled full size (run-off) and terminated RNA products from a reconstituted single round transcription reaction (see Materials and Methods). T, terminated transcript; %T, efficiency of termination; TPP, thiamin pyrophosphate.

In light of the limitations of energy minimization, this method must be used carefully, with certain caveats in mind. First, the aptamer domain of the bacterial riboswitch used as a reference must be well-predicted by energy minimization relative to the experimental structure. For example, the secondary structure of the aptamer of the purine riboswitch can successfully be predicted by either mfold⁸ or RNAfold¹¹ and it is anticipated that in future, more aptamers will have this virtue. Incidentally, sequence based methods²⁴ are more heavily strained when applied to analyses of the bacterial consensus sequence of the purine riboswitch than when used on other aptamers, and as a consequence, for specific cases, the structure based method proposed in³⁷ and refined in³⁵ is capable of a more robust analysis. Second, because energy minimization methods entail expensive run times on the order of $O(n^3)$, where n is the size of the sequence, a pre-processing step, in which genes that participate in the metabolic cycle associated with the riboswitch being investigated are collected, can be performed beforehand on the entire genome of interest. Instituting such a preliminary step can result in a considerable reduction in the number of subsequent computations.

Additional strategies are available for improving the efficiency of this structure based method. Incorporating sequence constraints, if available, such as in the case of riboswitches, is an essential component of the method. It can be done after the main structure based filtering step has reduced the number of candidates considerably. Each candidate can then be inspected manually as in³⁷ or in conjunction with the structure based search as suggested in³⁵ **Figure 4**, which was generated with UNAFold,⁹ displays an attractive candidate that was found in³⁷. Even though the in-line probing experiment that was performed to test its response to a change in guanine concentration did not yield the desired response, which is also not proposed in any way to be considered by the structural method at this time, it is evident that the method can yield biologically interesting candidates that cannot be achieved by sequence based methods alone.

Discussion

The results chosen for discussion were carefully selected to contain predictions that are backed by experiments. For continuity in presentation, the last topic from the Results section above will be discussed first.

Detection of novel riboswitches. The use of energy minimization methods, a robust procedure for finding new riboswitches originally formulated with the addition of computational geometry techniques to measure structure similarity,⁵⁴ has been further developed in³⁷ and refined in³⁵ resulting in a technique that is complementary to the existing methods. The specific method referred to here is structure based (with flexibility, allowing the insertion of sequence constraints). Practically speaking, the biological motivation behind the research was to formulate a method that will perform well where others fail, especially when attempting to find novel riboswitches in eukaryotes that are evolutionarily far from prokaryotes (compared with finding novel riboswitches in prokaryotes based on known riboswitches in prokaryotes), since evolution may have preserved structure but not sequence.

In using the energy minimization based method to search for a novel purine riboswitch in *Arabidopsis*, we found an attractive candidate that preserves not only structure, but also a significant number of nucleotides in the sequence (**Fig. 4**). Subsequent in-line probing experiments proved that our secondary structure prediction was correct but failed to show that the candidate binds guanine as a riboswitch does, as such interaction is currently not possible to detect using bioinformatics techniques. In future, the accumulation of ligand-binding information may generate bioinformatics methods capable of considering these interactions. From the evidence so far and based on the absence of any substantial novel eukaryotic riboswitches being detected since 2003, it seems that if there are more riboswitches in eukaryotes, their sequences and structure compositions may differ from those of the prokaryotes, and therefore, finding them remains a difficult challenge that will undoubtedly motivate additional work in this field.

The structure based method discussed here is also applicable to newly sequenced genes or to new classes of riboswitches discovered in prokaryotes, besides purine, for which the energy minimization prediction of aptamer structure conforms with the experimental result. Advanced methods, which do not search for a consensus reference sequence and structure but that must meet a concise list of properties, possibly including a moving window, may be used instead. The goal of creating bioinformatics techniques to find novel eukaryotic riboswitches remains a formidable challenge.

Rational design of riboswitches. Our first aim was to predict an anti-termination mutation in the TPP riboswitch. In *Bacillus subtilis*, the riboswitch shown in **Figure 2A** binds TPP, which triggers the formation of the terminator hairpin and the premature stoppage of transcription. In the absence of TPP, the anti-terminator structure that suppresses transcription termination emerges. Termination and anti-termination are two highly stable states with an energy barrier between them. In² it was established that the energy barrier can be surpassed by the binding of TPP,

but it is of considerable interest to investigate whether other events can also provide the means for crossing the energy barrier. Thus, the question arises whether point mutations in the riboswitch structure are capable of crossing back and forth across the energy barrier between the termination and anti-termination states (Fig. 2) without a change in TPP.

Several mutations were made in the investigated conserved sensor domain,² and the resulting termination efficiency was experimentally measured. Using mfold⁸ to initially predict the secondary structure of the wild type structure depicted in Figure 2A and then to computationally examine changes in the riboswitch in response to several randomly selected point mutations shows clearly that while other parts of the riboswitch responded with structural changes, the terminator hairpin remained unaltered. Thus, it is expected that at best, only a limited number of point mutations will be able to cause a switch from the terminator structure to an antiterminator structure. Using the approach detailed in²⁸ and described briefly in Materials and Methods, the unique point mutation A158U, which transforms the structure from the termination state to an anti-termination state, can be predicted. Indeed, using standard energy minimization methods, this mutation leads to a predicted antiterminator structure, as depicted in Figure 2B. It should be mentioned that by the method of correlated mutation analysis of Horowitz and coworkers,^{55,56} Noivirt and Horowitz adapted their correlated mutation analysis to nucleic acids and were able to discover independently the unique effect of a mutation in the aforementioned position. In the context of mutations, more than one approach identifies position 158 in the TPP-riboswitch as distinctive.

Our second aim was to predict a compensatory mutation to deleterious mutation A158U. Similar to the procedure for predicting the deleterious mutation, a compensatory mutation prediction was performed to find point mutations that restore the riboswitch to its original termination state. An “eigenvalue table” analogous to the one in²⁸ was generated, and 17 point mutations were found that switch an antiterminator structure to a terminator hairpin structure. All these mutations are located further upstream from the terminator position in the sequence, either closer to or within the thi-box. Mutation A28U was chosen for validation purposes because its location in the sequence is the furthest compensatory mutation from deleterious mutation A158U. This makes point mutation A28U the most peculiarly predicted mutation to restore back termination.

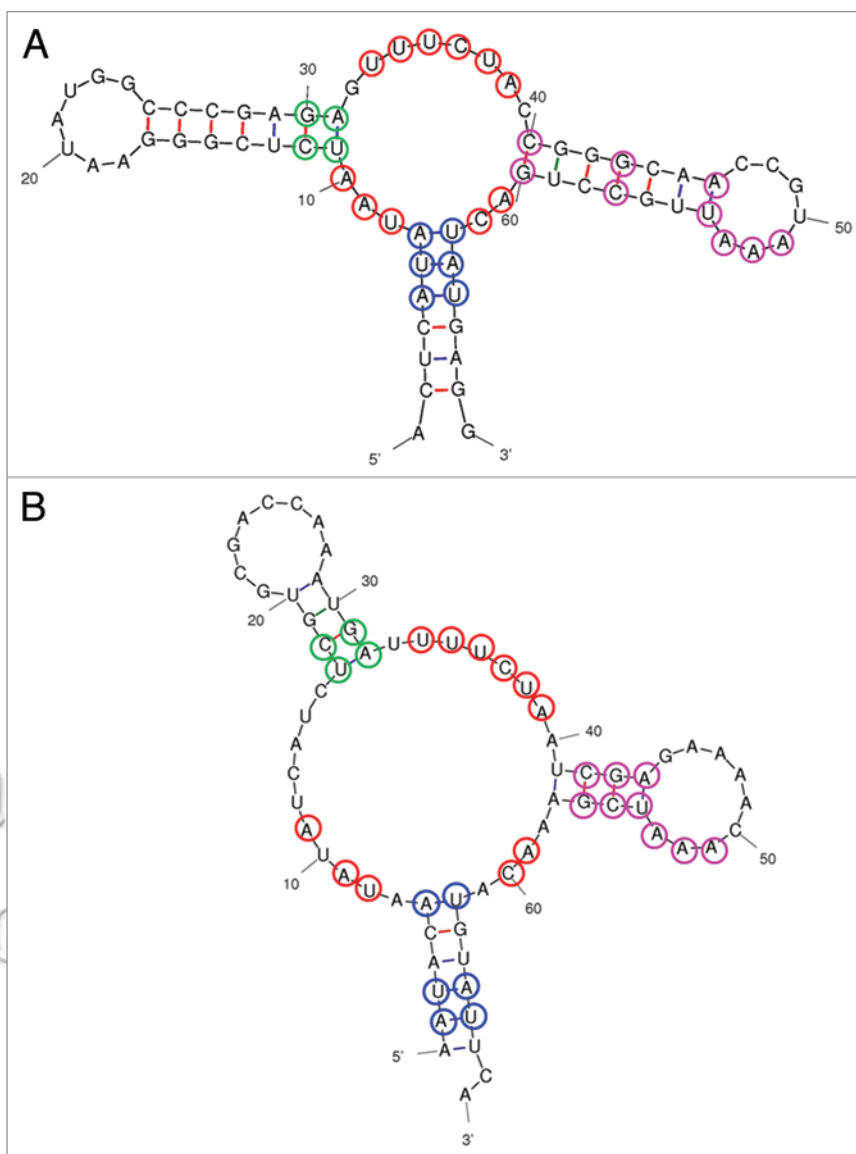


Figure 4. (A) The consensus purine riboswitch as available in RFAM. (B) A 3-way junction prediction finding using energy minimization³⁷ in the 3' region of the Arabidopsis *At2g05140* gene putatively encoding a phosphoribosylaminoimidazole carboxylase. The locations of sequence similarities between the consensus and our 3-way junction finding are marked with circles. All four nucleotides that participated in the hypoxanthine binding⁵⁰⁻⁵³ are present in our 3-way junction finding: U22 (of the last reference), U47, U51 and C74. Most of the common nucleotides appear in the core of the 3-way junction. Figure reproduced in part from.³⁷ UNAFold⁹ version 3.6 with mfold utils version 4.5 was used to predict and generate the secondary structure drawings.

Our third aim was to experimentally validate the predictions of both the deleterious and the restoring mutations. Using the experimental procedures of in vitro transcription (see Materials and Methods), three mutations were tested: A28U, A158U and U163A (for negative control). Mutation U163A was closer to the terminator, and there was no intuitive reason apart from the computational prediction to assume that A158U would cause a termination deficiency while U163A would not. Indeed, in support of that prediction, a decrease was observed in the termination efficiency for A158U at 25°C, and the compensatory mutation

A28U completely restored the termination deficient phenotype of A158U. With mutation U163A, no decrease in termination efficiency was observed, and the compensatory mutation A28U had no effect on the non-affected phenotype of U163A. None of the mutations altered the effect of TPP on the riboswitch.

It should be noted that all these experiments were performed at 25° and 38°C. At the latter temperature, no effect of any of the mutations on transcription termination was observed, whereas all of our findings are reported for the former temperature with the *Bacillus subtilis*. The predictions with mfold and the latest available energy parameters were performed with the default standard conditions. By itself, it was impressive that energy minimization methods succeeded in roughly capturing the prediction of a deleterious mutation A158U relative to a non-deleterious one (U163A) and the restoration as a consequence of A28U positioned 130 nt upstream. Thus, we can conclude that the non-obvious effects of the predicted mutations A158U and A28U were experimentally observed. We note that for RNA sequences that are approximately 200 nt long, the size of the TPP-riboswitch, it is expected that energy minimization methods will not accurately predict all base pairings because of the limitations of those methods. However, we argue that dramatic events like conformational switching can be captured by energy minimization methods, and there is an advantage to using them for carrying bioinformatics predictions that may be tested experimentally. Regarding our mutation predictions in the regulatory RNA system, such results may enable the artificial modification of a natural riboswitch found in vivo by introducing favorable point mutations for a variety of control purposes with potential applications in RNA gene regulation. This may also contribute in future research to biochemical efforts toward the design of engineered riboswitches.⁴¹

Materials and Methods

The study of RNA secondary structure prediction dates back to early works by Waterman⁵⁷ and the Nussinov-Jacobson algorithm⁵⁸ that use dynamic programming. Since the early 80's,⁵⁹ stacking energies have been incorporated into secondary structure prediction, and the energies that are minimized by the recursion are derived from empirical calorimetric experiments, such as the ones published in.⁴⁹ Zuker's mfold and the Vienna RNA package that are used for the predictions described here all rely on energy minimization methods using thermodynamics. At present, such predictive calculations are limited to RNA secondary structures, whereas in the future it may be possible to extend these predictions to include tertiary structure information when considering approaches and computational applications such as in.⁶⁰⁻⁶³ Nevertheless, due to the hierarchical nature of RNA folding,^{64,65} the secondary structure is indicative of the folding to tertiary structure. Moreover, a substantial amount of information can be inferred about the riboswitch mechanism by examining its secondary structure, since certain well-known secondary structure motifs that constitute riboswitches (e.g., the terminators/anti-terminators and sequestors/anti-sequestors) are known to possess functionally important roles. Although energy minimization methods are usually limited in their reliability to

predict the exact base pairings in RNA sequences longer than 100–150 nt, our mutation predictions assume that dramatic events at the coarse-grain level, such as a transition from a terminator structure to an antiterminator structure, can be captured successfully by these methods.

The effect of single point mutations on RNA secondary structure has been addressed before using an RNA tree-graph representation.⁶⁶ In²⁸ the problem was revisited in detail with riboswitch examples. A similarity measure was proposed that facilitates computational predictions by using spectral decomposition of the Laplacian matrix for the RNA tree-graphs,⁶⁷ gaining intuition regarding the compactness of those tree-graphs via mathematical theorems from spectral graph theory.^{68,69} For a detailed description and illustrative examples of the prediction method, see.^{28,40} The coarse-grained tree graph representation of RNA secondary structure was introduced in^{70,71} and was subsequently applied in⁶⁶ to examine the prediction of single point mutations in the L11 mRNA of *Escherichia coli*. Riboswitch mechanisms were not known at the time, and folding predictions generated using energy minimization has significantly improved since then.^{9,10,39,49,72} The introduction of spectral decomposition methods^{28,67} and the clustering of tree graphs into eigenvalues and other similarity measures in a user-friendly manner using a computerized tool^{43,44} allows for easy data organization that enables the prediction of deleterious mutations that disrupt desired motifs in riboswitches.

An equivalent way of representing an RNA tree-graph is by using a special matrix called the Laplacian. The Laplacian matrix corresponding to a graph is symmetric, with one row and column for each node on the graph, and its rows and columns sum to zero. It is constructed as follows: in the diagonal of the matrix, the degree of the vertex (number of incident edges) is specified, while in the off-diagonals the value “-1” is inserted if there is an edge at that location, or “0” if there is no connecting edge. The complete set of eigenvalues of the Laplacian matrix is called the spectrum of the graph, and is independent of how graph vertices are labeled. The following properties characterize the Laplacian matrix eigenvalues: (A) The eigenvalues of the matrix are nonnegative and the first eigenvalue is zero; (B) The second smallest eigenvalue is the algebraic connectivity of the graph;⁶⁸ (C) For a “star-shaped” tree-graph, the second smallest eigenvalue is unity.⁶⁹

It was originally proposed in,²⁸ based on analogies with domain decomposition in parallel computing, to consider the second eigenvalue of the Laplacian matrix as a measure of RNA tree-graph compactness. The second eigenvalue of the Laplacian matrix corresponding to a general graph is the measure of its graph connectivity. Intuitively, graph connectivity is monotonically increasing from its lowest value for a linear tree-graph structure to its highest value for a “star shaped” tree-graph structure. For tree-graphs, its value is between 0.0 and 1.0.⁴⁰ Other graph motif measures besides the eigenvalues can be used for the deleterious mutation prediction, but the theorems by Fiedler and Merris^{68,69} associating the second smallest eigenvalue of the Laplacian matrix to the algebraic connectivity of a tree-graph, facilitate intuitive organization of the data. Alternatively, the use of other similarity

measures, as was implemented in,^{43,44} can offer some improvements to the process of detecting deleterious mutations.

The proposed computational method is independent of the folding algorithms, but relies in the predictions presented here on Zuker's mfold⁸ and the Vienna RNA package,^{10,11} both of which used the energy parameters described in⁴⁹ and returned similar results in the case presented. Our method consists of starting from an initial prediction of the secondary structure of the wild type, and for each point mutation in the wild type sequence, we predict the secondary structure of the associated mutant and assess its similarity to the wild type structure. The second eigenvalues of the Laplacian matrix corresponding to the tree-graph representation of the secondary structure are used for measuring similarity between the mutants and the wild type. Thus, the results are organized in an "eigenvalue table" similar to that in.²⁸

Experimental procedures for in vitro transcription were used to validate our predictions. Point mutations in the thi-box leader were introduced by PCR-directed mutagenesis of the original template.² All transcription templates were generated by PCR and purified from low melting agarose and diluted in TE to 2 pmol/ μ l. The *E. coli* His6- β '-tagged RNAP σ ⁷⁰ holoenzyme was purified as described in.⁷³ The start-up elongation complex on the thi-leader templates was prepared as follows: 2 pmol DNA (1 μ l) and 1 pmol RNAP (1 μ l) were mixed in 10 μ l of the transcription buffer [TB: 100 mM KCl, 3 mM MgCl₂, 50 mM tris-HCl (pH 7.5)] and incubated 5 min at 37°C. Next, 1 μ l of the starting cocktail was added to give the final concentration of GpGpU primer of 20 μ M, ATP and GTP of 30 μ M, and [α -³²P] CTP (3,000 Ci/mmol) of 0.3 μ M. After 8 min at 37°C, the sample was diluted with TB and the aliquots were transferred to tubes containing all four NTPs (250 μ M final concentration). The chase reactions were stopped after ~15 min at 25°C by adding an equal volume of the sequencing gel loading buffer containing EDTA (200 mM) and formamide (95%) to each tube. Positions

of the terminated products were determined by the transcription sequencing reaction with 3'-dNTPs (TriLink). Relative amounts of [³²P] RNA were determined using PhosphorImager and software from Molecular Dynamics. The efficiency of termination (%T) was calculated by dividing the amount of radioactivity in a particular terminated band by the total radioactivity present in that and all read-through bands.

The effect of *thi* leader mutations on transcription was monitored by a single round run-off assay using PCR-generated linear DNA templates containing *B. subtilis tenA* promoter followed by a complete *thi*-leader sequence and the highly pure RNA polymerase holoenzyme from *E. coli*. In each case, the RNA transcript was [³²P] labeled near its 5' end during formation of the initial elongation complex (EC), which was stalled (see above) and then chased to the *thi*-leader terminator by adding all four substrates (NTP) at 25°C. Under these conditions, transcription was partially terminated at position +187 within the U-stretch of the attenuator (Fig. 3). The efficiency of termination (%T) was ~55% for the wild type template (Fig. 3, lanes 1) and increased to >95% in the presence of TPP to 50 μ M (lane 2). Mutation A158U decreased %T to 17%. Such low termination efficiency could be achieved at the wildtype template only at higher temperatures (37°C and above). Remarkably, the A28U mutation completely reversed the inhibitory effect of A158U on termination and increased %T back to ~55% (lane 7). Notably, A28U by itself had no effect on %T (lane 3). Neither mutations changed riboswitch responsiveness to TPP. Taken together, these data show that A28U is a highly specific compensatory mutation that acts at a large distance from the original deleterious mutation in the terminator.

Acknowledgements

This work was supported by a grant from the Israel USA Binational Science Foundation (BSF grant 2003291).

References

- Nahvi A, Sudarsan N, Ebert MS, Zou X, Brown KL, Breaker RR. Genetic control by a metabolite binding mRNA. *Chem Biol* 2002; 9:1043.
- Mironov AS, Gusarov I, Rafikov R, Lopez LE, Shatalin K, Kreneva RA, et al. Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell* 2002; 111:747-56.
- Winkler WC, Nahvi A, Breaker RR. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* 2002; 419:952-6.
- Winkler WC, Cohen-Chalamish S, Breaker RR. An mRNA structure that controls gene expression by binding FMN. *Proc Natl Acad Sci USA* 2002; 99:15908-13.
- Gelfand MS, Mironov AA, Jomantas J, Kozlov YI, Perumov DA. A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes. *Trends Genet* 1999; 15:439-42.
- Grundy FJ, Henkin TM. The S box regulon: a new global transcription termination control system for methionine and cysteine biosynthesis genes in gram-positive bacteria. *Mol Microbiol* 1998; 30:737-49.
- Miranda-Ríos J, Navarro M, Soberón M. A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteri. *Proc Natl Acad Sci USA* 2001; 98:9736-41.
- Zuker M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 2003; 31:3406-15.
- Markham NR, Zuker M. UNAFold: software for nucleic acid folding and hybridization. *Methods Mol Biol* 2008; 453:3-31.
- Hofacker IL, Fontana W, Stadler PF, Bonhoeffer LS, Tacker M, Schuster P. Fast folding and comparison of RNA secondary structures. *Monatsh Chem* 1994; 125:167-88.
- Hofacker IL. Vienna RNA secondary structure server. *Nucleic Acids Res* 2003; 31:3429-31.
- Sudarsan N, Barrick JE, Breaker RR. Metabolite-binding RNA domains are present in the genes of eukaryotes. *RNA* 2003; 9:644-7.
- Barrick JE, Corbino KA, Winkler WC, Nahvi A, Mandal M, Collins J, et al. New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc Natl Acad Sci USA* 2004; 101:6421-6.
- Barrick JE. Predicting riboswitch regulation on a genomic scale. *Methods Mol Biol* 2009; 540:1-13.
- Weinberg Z, Ruzzo WL. Exploiting conserved structure for faster annotation of non-coding RNAs without loss of accuracy. *Bioinformatics* 2004; 20:334-41.
- Yao Z, Weinberg Z, Ruzzo WL. CMfinder—a covariance model based RNA motif finding algorithm. *Bioinformatics* 2006; 22:445-52.
- Mandal M, Lee M, Barrick JE, Weinberg Z, Emilsson GM, Ruzzo WL, Breaker RR. A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* 2004; 306:275-9.
- Weinberg Z, Barrick JE, Yao Z, Roth A, Kim JN, Gore J, et al. Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res* 2007; 35:4809-19.
- Weinberg Z, Regulski EE, Hammond MC, Barrick JE, Yao Z, Ruzzo WL, Breaker RR. The aptamer core of SAM-IV riboswitches mimics the ligand-binding site of SAM-I riboswitches. *RNA* 2008; 14:822-8.
- Griffiths-Jones S, Bateman A, Marshall M, Khanna A, Eddy SR. Rfam: an RNA family database. *Nucleic Acids Res* 2003; 31:439-41.
- Gardner PP, Daub J, Tate JG, Nawrocki EP, Kolbe DL, Lindgreen S, et al. Rfam: updates to the RNA family database. *Nucleic Acids Res* 2009; 37:136-40.
- Kazanov MD, Vitreschak AG, Gelfand MS. Abundance and functional diversity of riboswitches in microbial communities. *BMC Genomics* 2007; 8:347.
- Bengert P, Dandekar T. Riboswitch finder—a tool for identification of riboswitch RNAs. *Nucleic Acids Res* 2004; 32:154-9.

24. Zhang S, Borovok I, Aharonowitz Y, Sharan R, Bafna V. A sequence-based filtering method for ncRNA identification and its application to searching for riboswitch elements. *Bioinformatics* 2006; 22:557-65.
25. Klein RJ, Eddy SR. RSEARCH: Finding homologs of single structured RNA sequences. *BMC Bioinformatics* 2003; 4:44.
26. Hammann C, Westhof E. Searching genomes for ribozymes and riboswitches. *Genome Biol* 2007; 8:210.
27. Epshtein V, Mironov AS, Nudler E. The riboswitch-mediated control of sulfur metabolism in bacteria. *Proc Natl Acad Sci USA* 2003; 100:5052-6.
28. Barash D. Deleterious mutation prediction in the secondary structure of RNAs. *Nucleic Acid Res* 2003; 31:6578-84.
29. Bompfünower AF, Flamm C, Fried C, Fritsch G, Hofacker IL, Lehmann J, et al. Evolutionary patterns of non-coding RNAs. *Theory Biosci* 2005; 123:301-69.
30. Steffen P, Voss B, Rehmsmeier M, Reeder J, Giegerich R. RNASHapes: an integrated RNA analysis package based on abstract shapes. *Bioinformatics* 2006; 22:500-3.
31. Freyhult E, Moulton V, Clote P. Boltzmann probability of RNA structural neighbors and riboswitch detection. *Bioinformatics* 2007; 23:2054-62.
32. Freyhult E, Moulton V, Clote P. RNABor: a web server for RNA structural neighbors. *Nucleic Acids Res* 2007; 35:305-9.
33. Busch A, Backofen R. Info-RNA—a server for fast inverse RNA folding satisfying sequence constraints. *Nucleic Acids Res* 2007; 35:310-3.
34. Avihoo A, Gabdank I, Shapira M, Barash D. In Silico Design of Small RNA Switches. *IEEE Transactions on Nanobioscience Special Issue on Computational Nanobioscience* 2007; 6:4-11.
35. Vekslar-Lublinsky I, Barash D, Ziv-Ukelson M, Kedem K. *J Comp Biol* 2007; 14:908-26.
36. Barash D, Sikorski J, Perry L, Nevo E, Nudler E. Adaptive mutations in RNA-based regulatory mechanisms: computational and experimental investigations. *Israel J Ecol Evol* 2007; 52:263-79.
37. Cohen A, Bocobza S, Vekslar-Lublinsky I, Gabdank I, Barash D, Aharoni A, et al. Computational identification of three-way junctions in folded RNAs: a case study in arabidopsis. *In Silico Biology* 2008; 8:105-20.
38. Geis M, Flamm C, Wolfinger MT, Tanzer A, Hofacker IL, Middendorf M, et al. Folding kinetics of large RNAs. *J Mol Biol* 2008; 379:160-73.
39. Gruber AR, Lorenz R, Bernhart SH, Neuböck R, Hofacker IL. The Vienna RNA Websuite. *Nucleic Acids Res* 2008; 36:70-4.
40. Barash D. Second eigenvalue of the Laplacian matrix for predicting RNA conformational switch by mutation. *Bioinformatics* 2004; 20:1861-9.
41. Bauer G, Suess B. Engineered riboswitches as novel tools in molecular biology. *Journal of biotechnology* 2006; 124:4-11.
42. Weigand JE, Sanchez M, Gunnesch EB, Zeiher S, Schroeder R, Seuss B. Screening for engineered neomycin riboswitches that control translation initiation. *RNA* 2008; 14:89-97.
43. Churkin A, Barash D. RNAmute: RNA secondary structure mutation analysis tool. *BMC Bioinformatics* 2006; 7:221.
44. 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.
45. Shu W, Bo X, Liu R, Zhao D, Zheng Z, Wang S. RDMAS: a web server for RNA deleterious mutation analysis. *BMC Bioinformatics* 2006; 7:404.
46. Waldspühl J, Devadas S, Berger S, Clote P. Efficient algorithms for probing the RNA mutation landscape. *PLoS Comput Biol* 2008; 4:1000124.
47. Waldspühl J, Devadas S, Berger S, Clote P. RNAmutants: a web server to explore the mutational landscape of RNA secondary structures. *Nucleic Acids Res* 2009; 37:281-6.
48. Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS. Comparative genomics of thiamin biosynthesis in prokaryotes. New genes and regulatory mechanisms. *J Biol Chem* 2002; 277:48949-59.
49. Mathews DH, Sabina J, Zuker M, Turner DH. Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* 1999; 288:911-40.
50. Batey RT, Gilbert SD, Montange RK. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* 2004; 432:411-5.
51. Serganov A, Yuan YR, Pikovskaya O, Polonskaia A, Malinina L, Phan AT, et al. Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem Biol* 2004; 11:1729-41.
52. Holbrook SR. RNA structure: the long and short of it. *Curr Opin Struct Biol* 2005; 15:302-8.
53. Lescoute A, Westhof E. Riboswitch structures: purine ligands replace tertiary contacts. *Chem Biol* 2005; 12:10-3.
54. Bergig O, Barash D, Nudler E, Kedem K. STR2: A structure to string approach for locating G-box riboswitches in pre-selected genes. *In Silico Biology* 2004; 4:593-604.
55. Kass I, Horovitz A. Mapping pathways of allosteric communication in GroEL by analysis of correlated mutations. *Proteins: Struct Funct Genet* 2002; 48:611-7.
56. Noivirt O, Eisenstein M, Horovitz A. Detection and reduction of evolutionary noise in correlated mutation analysis. *Protein Eng Des Sel* 2005; 18:247-53.
57. Smith TF, Waterman MS. RNA secondary structure: a complete mathematical analysis. *Math Biosci* 1978; 42:257-66.
58. Nussinov R, Jacobson AB. Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proc Natl Acad Sci* 1980; 77:6309-13.
59. Zuker M, Stiegler P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res* 1981; 9:133-48.
60. Parisien M, Major F. The MC-Fold and MC-Sym pipeline infers RNA structure from sequence data. *Nature* 2008; 452:51-5.
61. Martinez HM, Maizel JV, Shapiro BA. RNA2D3D: A program for generating, viewing, and comparing 3-dimensional models of RNA. *J Biomol Struct Dyn* 2008; 25:669-83.
62. Jonikas MA, Radmer RJ, Laederach A, Das R, Pearlman S, Herschlag D, Altman RB. Coarse-grained modeling of large RNA molecules with knowledge-based potentials and structural filters. *RNA* 2009; 15:189-99.
63. Dror O, Nussinov R, Wolfson HJ. The ARTS web server for aligning RNA tertiary structures. *Nucleic Acids Res* 2006; 34:412-5.
64. Brion P, Westhof E. Hierarchy and dynamics of RNA folding. *Annu Rev Biophys Biomol Struct* 1997; 26:113-37.
65. Tinoco I, Bustamante C. How RNA folds. *J Mol Biol* 1999; 293:271-81.
66. Margalit H, Shapiro BA, Oppenheim AB, Maizel JV. Detection of common motifs in RNA secondary structures. *Nucleic Acids Res* 1989; 17:4829-45.
67. Barash D. Spectral decomposition for the search and analysis of RNA secondary structure. *J Comp Biol* 2004; 11:1169-74.
68. Fiedler M. Algebraic connectivity of graphs. *Czechoslovak Math J* 1973; 23:298-305.
69. Merris R. Characteristic vertices of tress. *Lin Mult Alg* 1987; 22:115-31.
70. Shapiro BA. An algorithm for comparing multiple RNA secondary structures. *Comput Appl Biosci* 1988; 4:387-93.
71. Le SY, Nussinov R, Maizel JV. Tree graphs of RNA secondary structures and their comparisons. *Comput Biomed Res* 1989; 22:461-73.
72. Zuker M. Calculating nucleic acid secondary structure. *Curr Opin Struct Biol* 2000; 10:303-10.
73. Nudler E, Avetisova E, Markovtsov V, Goldfarb A. Transcription processivity: protein-DNA interactions holding together the elongation complex. *Science* 1996; 273:211-7.