In Silico Design of Small RNA Switches

Assaf Avihoo, Idan Gabdank, Michal Shapira, and Danny Barash*, *Member, IEEE*

***Abstract*—The discovery of natural RNA sensors that respond to a change in the environment by a conformational switch can be utilized for various biotechnological and nanobiotechnological advances. One class of RNA sensors is the riboswitch: an RNA genetic control element that is capable of sensing small molecules, responding to a deviation in ligand concentration with a structural change. Riboswitches are modularly built from smaller components. Computational methods can potentially be utilized in assembling these building block components and offering improvements in the biochemical design process. We describe a computational procedure to design RNA switches from building blocks with favorable properties. To achieve maximal throughput for genetic control purposes, future designer RNA switches can be assembled based on a computerized preprocessing buildup of the constituent domains, namely the aptamer and the expression platform in the case of a synthetic riboswitch. Conformational switching is enabled by the RNA versatility to possess two highly stable states that are energetically close to each other but topologically distinct, separated by an energy barrier between them. Initially, computer simulations can produce a list of short sequences that switch between two conformers when trigerred by point mutations or temperature. The short sequences should possess an additional desirable property; when these selected small RNA switch segments are attached to various aptamers, the ligand binding mechanism should replace the aforementioned event triggers, which will no longer be effective for crossing the energy barrier. In the assembled RNA sequence, energy minimization folding predictions should then show no difference between the folded structure of the entire sequence relative to the folded structure of each of its constituents. Moreover, energy minimization methods applied on the entire sequence could aid at this preprocessing stage by exhibiting high mutational robustness to capture the stability of the formed hairpin in the expression platform. The above computer-assisted assembly procedure together with application specific considerations may further be tailored for therapeutic gene regulation.**

***Index Terms*—Design of RNA switches, energy minimization methods, RNA folding predictions.**

I. Introduction

THE RNA molecule, once considered as an intermediate step between DNA and proteins, has drawn much attention in recent years also due to its unique capabilities that can be utilized in biotechnology. Although the functional role of RNAs are often related to their 3-D structure, the RNA secondary structure is experimentally accessible and contains much information to shed light on the relationship between structure and function. In general, RNA folding is thought to be hierarchical in nature [7], [37] where a stable secondary structure forms first and subsequently there is a refinement to the tertiary fold. Thus, RNA conformational rearrangements can often be studied to first order by examining their secondary structure, while keeping in mind the importance of tertiary structure [15]. For example, in the newly discovered genetic control elements called riboswitches [18], [26], a mechanism for bacterial gene regulation can be observed by examining the secondary structure alone. A switch between two highly stable states occurs as a consequence of direct metabolite binding that allows it to cross the energy barrier between these two states. In addition, conformational switching in the secondary structure can also be achieved by other event triggers (see Fig. 1 for a schematic illustration). For example, it was noted in [31] that there is some probability that even a single mutation can substantially alter the RNA secondary structure. Experimentally, this was observed in the spliced leader of *Leptomonas collosoma* [16], among other biological systems. Reviews and valuable information about RNA secondary structure prediction [45] as well as RNA switches with event triggers other than metabolite binding are available in [25] and [27]. Before their discovery as RNA-based regulatory mechanism used in natural systems, Breaker and coworkers were attempting to biochemically design artificial RNA switches that respond to metabolite binding [35] in laboratory experiments (a recent review is available in [6]).

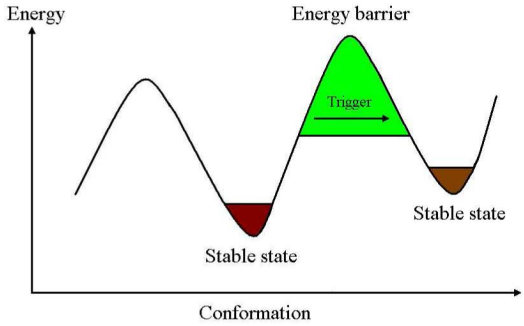


Fig. 1. Schematic illustration of a simple RNA switch.

On the prediction side, given an RNA sequence, the paRNAss algorithm has been developed to evaluate the predictability of conformational switching in that sequence [39]. However, paRNAss does not use temperature and mutations as variables in the prediction, nor any other event triggers from the external environment and therefore will not categorize several known riboswitches as well as the small RNA candidate sequences reported in [1] as being RNA switches. When attempting to computationally design new artificial switches, the seminal

Manuscript received June 30, 2006; revised October 2, 2006. This work was supported in part by the Israel USA binational science foundation under Grant BSF 2003291. *Asterisk indicates corresponding author.*

A. Avihoo is with the Department of Computer Science, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel.

I. Gabdank is with the Department of Computer Science and the Department of Life Sciences, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel.

M. Shapira is with the Department of Life Sciences, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel (e-mail: shapiram@bgumail.bgu.ac.il).

*D. Barash is with the Department of Computer Science, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel (e-mail: dbarash@cs.bgu.ac.il).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNB.2007.891894

work of [11] can prove useful, since a multistable RNA is effectively a switch when the exact event triggers to cross the energy barrier between the stable states are found. In [29], an interesting computational strategy was devised to design artificial RNA constructs for the ligand-binding domain of the riboswitches (or aptamer) by simulating an effector (or trigger) using the McCaskill partition function algorithm [22] and subsequent work [5]. All of these approaches utilize RNA secondary structure predictions in their core. Another such design approach using energy minimization methods, unrelated to conformational switching, has been devised in [8] for designing the most stable and unstable mRNA sequences which code for a target protein. Here, we continue the path of simulating conformational switching on a computer using energy minimization methods and our goal is to computationally design RNA switches from building blocks. Three building blocks in the expression platform domain (in contrast to the ligand-binding domain that is worked out in [29], for expression platform and related nomenclature see [18]) have been suggested in [1]. To expand this line with an overview on the entire riboswitch and its constituent domains, more specific expression platform building blocks should be predicted in order to fulfill the constraint that attached to an aptamer, the switching ability by an external trigger will most likely be retained. Thus, the modular nature of riboswitches [18] motivates the predictions reported here with.

After briefly describing the four different similarity measures used in [1] to predict initial RNA switch building blocks, we first examine the feasibility of predicting and experimentally verifying a small RNA temperature switch. The candidate temperature switch is taken from [1] and an inline probing experiment [36], [42] is performed for conformational switching validation. We observe that while there is experimental agreement with the candidate RNA temperature switch, the issue of temperature is still difficult in terms of accurate predictions. Thus, we come to a conclusion that predictions of single-point mutation switches are better suited in our overall approach where we wish to computationally design a small switch segment that will preserve switching properties when attached to an aptamer. Next, we obtain novel building blocks that are computationally optimized to possess switching property and retain their structure when a G-box is attached to them. At the core of our design strategy, we use energy minimization folding predictions on the corresponding sequences with *mfold* [46] or the Vienna package [14]. The latest available energy parameters [21] are used. First, we describe the pattern representations and similarity measures that are used for the computational predictions, where the representations consist of both coarse-grained and fine-grained graphs that correspond to the RNA secondary structure. We summarize some computational methods that were developed previously [1], [3] and are used in this article. For more details, the interested reader is referred to these references. Second, we describe some wet-lab experimental procedures that we have started implementing for verifying our computational predictions. To exemplify the interplay between computer predictions and experimental verifications (as was recently performed in [28]) in our approach, we provide some experimental results for one of our previously predicted temperature switch candidates [1]. Third, we expand on the issue of single-point mutation switch candidates and justify why these offer the easiest alternative for the prediction, coupled with experimental valdation. After providing a realistic test-case analysis of mutational robustness distribution showing computationally that RNA single-point mutation switches are relatively rare and hence peculiar, we choose a new synthetic single-point small RNA switch candidate for further illustration. We then consider how to construct an entire riboswitch with a sensor domain given a single-point mutation switch candidate. Finally, we report folding prediction simulations by attaching to the building block candidates a possible G-box domain [19], [26], [38] without loss of generality. In particular, the G-box was chosen for simplicity, though other aptamer domains can be inserted. It should be mentioned that the secondary structure of the G-box is well predicted by energy minimization methods in the presence of guanine, but the G-box changes its structure in the absence of guanine when considering the guanine riboswitch as a whole during the formation of the anti-terminator structure [4]. Thus, we are careful in our computational procedure not to consider the secondary structure prediction of the gunanine riboswitch in its anti-termination state. In the conclusion, we describe how our RNA building block candidates in the expression platform will be subjected to further laboratory experiments for the verification of their switching properties. We emphasize that unlike the building blocks reported in [1], the building block candidates predicted here are calculated specifically with the purpose of designing synthetic riboswitches when attached to ligand-binding domains.

II. Computational Methods

The information contained in the RNA secondary structure can be simplified to contain the most desirable features and hence be represented as a pattern. This can be achieved by using various representations such as graphs, strings, or matrices. In each representation type, there is the possibility to represent the secondary structure at the nucleotide level (fine grain representation) or at the motif level (coarse grain representation). To obtain a matrix representation equivalent to a graph, let $T = (V, E)$ be a tree-graph with vertex set $V = v_1, v_2, \ldots, v_n$ 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 $A(T)$ [9], [23]) is $L(T) = (a_{ij})$, where

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

$L(T)$ 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 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_n = 0$ the eigenvalues of $L(T)$. The second smallest eigenvalue, λ_{n-1}, is called the algebraic connectivity [9] of T and labeled as $a(T)$. Some properties of $a(T)$ that are relevant to the application presented here will be briefly mentioned in the next Section.

To illustrate a coarse-grain and a fine-grain representation of RNAs using matrices, let us examine the sequence CCCAACC-

CAAAGGGAAGGG. This short sequence will fold by *mfold* or *RNAfold* to yield the prediction "(((..(((. . .)))..)))" in dot bracket notation [13]. Thus, in a matrix representation at the level of nucleotides, the corresponding Laplacian matrix reads as shown at the bottom of the page (see [1] for other examples) and at the level of motifs, it should be observed that the coarse-grained simplification of "CCCAACCCAAAGGGAAGGG" contains the $5' - 3'$ end, a stem, an internal loop, another stem, and a hairpin composed of the three A's. Thus, the coarse-grain tree graph is a path of three vertices, where the middle vertex has two connected neighbors. Thus, at the coarse-grain level of motifs, the Laplacian matrix reads (see [3] for other examples)

$$L = \begin{pmatrix} 1 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}$$

when taking into account the $5' - 3'$ end, in addition to the internal loop and the hairpin, as vertices in the motif level tree-graph of a path of three vertices that the familiar Laplacian operator matrix arising from finite difference corresponds to. In addition, unrelated to the above example, we chose not to assign vertices to internal loops and bulges consisting of a single nucleotide on either side of the loop. Next, we describe several similarity measures that are used to estimate how two pattern representations differ at both the fine-grain and coarse-grain levels, as well as a hybrid between the two. Other interesting similarity measures are available in [34], [39], and [44].

In our current simulation, we use four different similarity measures (described briefly here; for more details see [1]). The first is the second eigenvalue of the Laplacian matrix corresponding to the coarse-grained tree-graph representation of the RNA secondary structure. Despite its expected limitations when dealing with small RNAs due to the low amount of secondary structure motifs, in this particular case it happens to be a good switch indicator using simple intuition because the coarse-grained tree graphs typically contain either two or three vertices. The second similarity measure we describe is the Wiener number of the Laplacian matrix corresponding to the fine-grained graph representation of the secondary structure, as suggested by Merris [24]. The third is the RNAdist in the Vienna package [14], relying on the fine-grained graph that is equivalent to the dot-bracket notation often used to represent the secondary structure [13]. The fourth is an in-house hybrid change tracker (HCT), utilizing coarse-grained definitions from [30] as letters in a string that is being generated at the nucleotide level. Below, we elaborate more on these four similarity measures.

The second eigenvalue of the Laplacian matrix, as described in [3], starts from Shapiro's coarse-grain representation [20], [33], as calculated using the "b2shapiro" routine available in [13]. An equivalent representation of the coarse-grained tree-graphs is a Laplacian matrix, for which its second eigenvalue is a measure of the tree-graph compactness [9]. When using the Laplacian matrix $L(T)$ to represent the tree T, the following mathematical properties become useful.

- The eigenvalues of $L(T)$ are nonnegative and the first eigenvalue is zero.
- The second smallest eigenvalue is the algebraic connectivity [9] of T, denoted by $a(T)$.
- $0 \leq a(T) \leq 1$.
- $a(T) = 2(1-cos(\pi/n))$ iff $T = P_n$ is a path on n vertices [9].
- $a(T) = 1$ iff $T = K_{1,n-1}$ is a star on n vertices [12], [23].

Interestingly, for the case of a star of any number of vertices N when $N > 2$, the second eigenvalue is 1.0 [23]. However, for the exceptional case when $N = 2$, the second eigenvalue is 2.0 (see [3], [23]) as occurs in the examples available in [1]. Thus, the difference between a star of two and three vertices

$$\mathrm{L} = \begin{pmatrix}
2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 \\
-1 & 3 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 \\
0 & -1 & 3 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\
0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 3 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 3 & -1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 3 & -1 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & -1 & 3 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & -1 & 3 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 3 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 \\
0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 3 & -1 & 0 & 0 \\
0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 3 & -1 & 0 \\
-1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 1
\end{pmatrix}$$

can be easily traced by the Laplacian second eigenvalue, which serves as a rough indicator for conformational transitions in the cases presented in [1]. In other cases, such a similarity measure may not capture all the details when encountering finer conformational transitions. Therefore, as proposed by Merris [24], the reciprocals of all eigenvalues of the Laplacian matrix for the fine-grained graph (equivalent to the dot-bracket representation) can be added to extract the Wiener index [41]. The Wiener index is a topological index defined as the sum of distances between all pairs of vertices in a tree. It has been used as a structural descriptor for molecular graphs. As proved in [24], the following corollary holds:

Let $T = (V, E)$ be a tree with vertex set $V = \{v_1, v_2, \ldots, v_3\}$ and Laplacian eigenvalues $\lambda_1 \geq \ldots \geq \lambda_{n-1} > \lambda_n = 0$. Then the Wiener index of T is

$$W(T) = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} d(v_i, v_j) = n \sum_{i=1}^{n-1} 1/\lambda_i$$

where $d(v_i, v_j)$ is a count of the edges in the unique path from v_i to v_j. This similarity measure, as well as the edit distance on strings provided by the RNAdist available in [13], both work at the level of fine-grained graph representations. We note that unlike information extracted from the Laplacian eigenvalues that often has an intuitive meaning but does not constitute a metric, the edit distance on strings is a metric. Finally, we also use an in-house HCT, a heuristic procedure that we developed specifically for our purposes and was described in [1]. It uses a hybrid strategy that combines coarse and fine grain representations. Coarse-graining can be mathematically defined in several ways and here we use the definitions found in [30] that distinguish between six different types of secondary structure motifs: single strand, double strand, bulge, hairpin, interior loop, and multibranched loop. Throughout our design procedure, we use in conjunction the aforementioned pattern representations, similarity measures, and energy minimization folding predictions [14], [46].

III. Experimental Methods

To begin our computational approach, we first examine the possibility to start from generating building blocks of small RNA temperature switches, validated experimentally. Specifically, we take candidate A from [1] and perform an experimental verification whether switching indeed occurs in two different temperature. Although our prediction in [1] corresponds to two adjacent temperatures, 36 °C and 37 °C, we increase the upper limit to 42 °C for the following reason emanating from the experimental setup. Suppose we have a solution that contains the RNA and is kept at room temperature (25 °C). For the inline probing experiment we take the solution, add a buffer, and put the RNA in a chamber of 36 °C. After some time we take the RNA out of the chamber and run it on a gel to obtain the inline probing result. If we repeat the same procedure with 37 °C (i.e., put the RNA in a chamber that is 37 °C), there is clearly a need to take into account some kind of an activation energy since the RNA already acquired a structure at room temperature (25 °C). When heating the RNA, its structure may likely not change unless some kind of an energy barrier is overcome. Thus, in addition to our experimental results that are described here with and may not seem to have a definite conclusion, the activation energy needs to be accurately predicted with a set of premeasured enthalpy parameters though the computational approach described in [10] and [43] is well fit for modeling energy barriers. This may introduce another complication when RNA temperature switch candidates are dealt with at present. Therefore, we have reached a conclusion that experimental verification and accurate computational predictions will be most conveniently performed at present with single-point mutation small RNA switch building blocks. Thus, we take as a consequence the small RNA mutation switch candidate predictions as our preferred procedure when resuming the computational simulations in Section V.

A. In Vitro Transcription

We begin the procedure for testing the small RNA switch candidate A from [1]. The small RNA switch candidate with the following sequence 5′-GGAUGAGCGUUACUGAAGAG-GAGGACCUCC-3′ was transcribed *in vitro*. A DNA template was produced by overlapping extension PCR. Two primers with a partial overlapping complementary sequence were extended to create a blunt ended template, which was further amplified with a second set of primers derived from the ends of the DNA template. Primer (a) from the first set consisted of the T7 promoter (17 nt) followed by 12 nt derived from the 5′ end of the small RNA switch (5′-TAATACGACTCACTATAG-GATGAGCGTTA-3′). Primer (b) of the first set consisted of 18 nt derived from the 3′ end of the small RNA switch, followed by 11 nt with complementary sequences to primer (a) (5′-GGAGGTCCTCCTCTTCAGTAACGCTCATC-3′). The final sequence was amplified by two nonoverlapping primers, primer (c) (5′-TAATACGACTCACTATAGGATGAG-3′) and primer (d), (5′-GGAGGTCCTCCTCTTCAGTAAC-3′). These primers were derived from the 5′ and the 3′ ends of the extended DNA template. The PCR product served as a template for *in vitro* transcription of the thermosensor RNA candidate with T7 RNA polymerase (Fermentas). The RNA was purified by size exclusion over a Sephadex G-50 column, dephosphorylated with calf-interstinal alkaline phosphatase (Fermentas) and radiolabeled at the ′ end with polynucleotide kinase (Fermentas). The radiolabeled RNA product was gel purified over 7M Urea-16% PAGE, and used for structure determination.

B. In-Line Probing

In-line probing reveals the structural context of nucleotides based on their relative susceptibility to an in-line attack of the $5' - 3'$ phosphodiesther bond of the RNA backbone from the adjacent $2'$ hydroxyl group of the ribose moiety in the RNA, making single-stranded regions more susceptible for cleavage. Radiolabeled RNA samples (∼10,000 cpm) were subjected to in-line probing by using the published protocols in [36], [42]. More specifically, radiolabeled RNA dissolved in water was incubated for 30 minutes at 35 °C and 42 °C. Following this step, the in-line probing buffer was added (to reach final concentrations of 20 mM MgCl2, 50 mM Tris : HCl pH 8.3 and 100 mM KCl) and the RNA was incubated for 6 hours at the corresponding temperatures. The partially cleaved RNA products

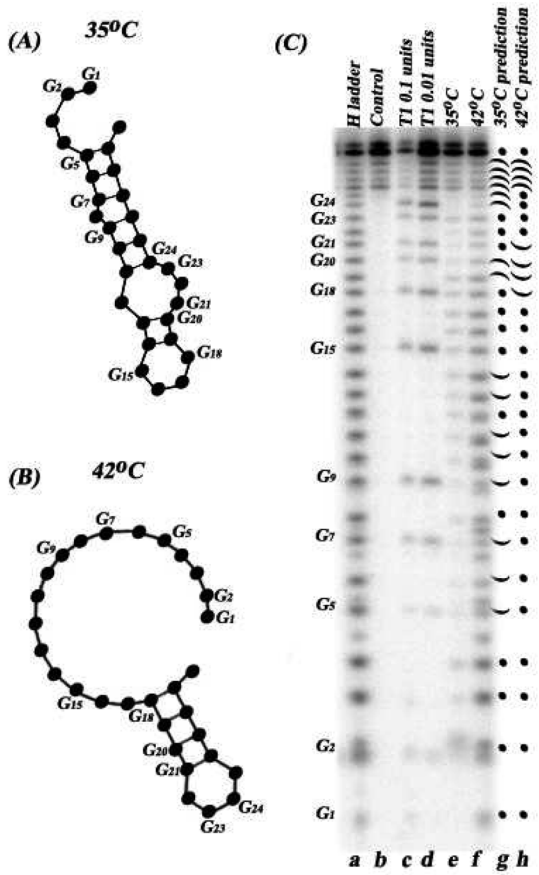


Fig. 2. Temperature induced structural changes in small RNA temperature switches. The secondary structure of a thermoswitch candidate was predicted by the Vienna package for (a) 35 °C and (b) 42 °C. These results were also verified with Zuker's mfold Version 2.3 and corresponding energy rules [40] (with enthalpies) where temperature is available as a variable. RNA encoding the predicted small RNA switch was synthesized *in vitro* using a PCR fragment as template. (c) The RNA was end labeled and subjected to in-line probing, for structure determination of the RNA at different temperatures. The partially cleaved RNA was analyzed on a 7M Urea-16% polyacrylamide gel (lanes e and f), parallel to an RNA size ladder with a single-nucleotide resolution (lane a), untreated RNA (lane b) and a G ladder which was obtained by a partial enzymatic digestion with RNase T1 (lanes c and d). A dot-bracket notation of the predicted secondary structure of the RNA at 35 °C and 42 °C is shown in lanes g and h, respectively.

were resolved on 7M Urea-16% PAGE that were dried and subjected to phosphorimager analysis. For the enzymatic probing, RNAse T1 reactions that produced a G-ladder of the radiolabeled RNA molecules were performed according to the manufacturer instructions (Ambion) with 0.1 and 0.01 units, and the products were analyzed over 7M Urea-16% PAGE. The results are available in Fig. 2.

C. Polyacrylamide Gel Electrophoresis (PAGE) of RNA Fragments

Electrophoresis through polyacrylamide gels is a standard method used to separate macromolecules, in our case fragments of nucleic acids, based on physical parameters such as size. RNA molecules are negatively charged due to the phosphate groups that comprise their backbone, and thus migrate towards the anode when an electric field is applied. The RNA fragments are separated over denaturing gels in the presence of 7M urea, to make sure that all fragments are linear, and that they do not form secondary structures that can change their pattern of separation. The different fragments migrate to different distances relative to their size, with small fragments migrating faster than larger ones, and since they are end-labeled with a radioactive phosphate group, they can be easily visualized on X-ray films, or by analysis with a Phosphorimager.

The "in-line probing" approach takes advantage of the instability of the RNA backbone under alkaline conditions. Prolonged incubations at room temperature under these conditions result in spontaneous cleavages at the backbone linkages preferentially in less stable sites, which are usually single stranded [35]. Separation of the cleavage products is done on 7 M urea 6%–8% polyacrylamide gel electrophoresis, along with size markers and H and G ladders (that represent cleavage which occurs statistically near each nucleotide, and near each G nucleotide, respectively).

IV. Experimental Results

In-line probing experiments confirm an overall occurrence of temperature dependent structural changes in the small RNA switch candidate that was tested though problematic issues appear. T_m of RNA varies for different sequences, and RNA structure is highly dependent on its temperature. Our calculations aim to design small RNA switches that change their structure at a given temperature. Our experiments verify the basic structure predicted at the lower temperature range, and monitor the conformational changes that take place upon temperature elevation by 7 °C. These changes are mostly in accordance to the predictions. In order to conveniently track these changes in the gel presented in Fig. 2(c), it is possible to examine the dot-bracket representation of the secondary structure prediction in both temperatures, available in lanes g and h on the right hand side of the gel. When reading the gel lanes e and f at temperatures 35 °C and 42 °C from the bottom up [it should be noted that G's on the left hand side of the gel correspond to the G's labeled on the secondary structure simplified representation in Fig. 2(a) and 2(b)], locations where in one temperature there is a single strand and in another there is a double strand and *vice versa* should be characterized by changes in the intensity of the bands in the gel. As can viewed in Fig. 2(a) and 2(b), the predicted structure involves an internal reorganization of the nucleotide base pairing. The stem strand that extends between G24 and C29 switches its pairing strand from G5 - U11 to G18-G21 and nucleotides G5-U11 switch from a double- to single-stranded form. As a result of this reorganization, intermediate states may be present in the solution, thus explaining some inconsistencies between the predicted and actual structure. Overall, though validation was mostly successful, the somewhat problematic interpretation of the results due to intermediate states and the issue of activation energy that is nonsufficient to predict at present in terms of accuracy made us suggest to continue with simulations of small RNA single-point mutation switches instead of small RNA temperature switches. This is our preferred procedure when performing our computational simulations described in the next Section. Other external factors, such as ligand binding, can also enhance switching abilities but we currently have no reliable way to

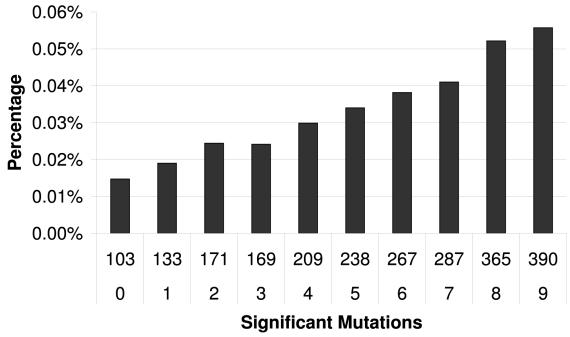


Fig. 3. Mutational robustness distribution for the initial pool of modular randomly generated sequences that contain a G-box domain. Percentage on the y axis is from the pool of 700,000 random nucleic acid sequence segments (see text) and significant mutations on the x axis is the number of mutations that lead to a conformational rearrangement of the secondary structure.

model them computationally. Advances in modeling such interactions in the future and case specific knowledge on the biology of the problem at hand may certainly contribute to our prediction abilities, but mutational robustness is currently the most computationally convenient measure to model for examining the relative stability of the structure and its capability to undergo a conformational rearrangement.

V. Computational Results

Choosing the single-point small RNA switch alternative over temperature switch candidates, we start by generating a random nucleic acid sequence pool of 700 000 segments with a length of 107 nt each in order to feasibly remain as much as possible within the scope of reliable predictions by energy minimization. Out of the 107 nt, 67 nt are constrained to be a G-box domain of one of the Bacilli reported in [19] where we took the *Bacillus subtilis-purE* without loss of generality (for the G-box, experimental results and structure determination is already available [4], [15], [17], [32]). The other 40 nt are randomly generated. At the initial stage, we perform a screening process and retain only those sequences in which the folding prediction of the G-box by itself and the folding prediction of the G-box piece within the entire sequence are identical. After this initial screening, we perform a histogram analysis (Fig. 3) retaining the sequences for which nine-point mutations or less will perturb the predicted secondary structure of a sequence, distributing the sequences accordingly. For the mutational robustness distribution, we use the Vienna package in conjunction with the pattern representation and similarity measures discussed in the previous section. We observe that only a small fraction of approximately 0.015% (103 sequences) will possess mutational robustness such that no single-point mutation will succeed to perturb the stable structure that is formed. The high mutational robustness of natural riboswitches was computationally observed in [2], and the fact that only a small fraction of sequences possesses this property is a further indication that such a screening procedure may yield successful synthetic riboswitches.

For the next stage, out of the 103 sequences that are of high mutational robustness according to Fig. 3, we take out the G-box domain and remain with only very few sequences in which a single-point mutation will succeed to perturb the small RNA segment of 40 nt contained in the expression platform region. In the final stage, we apply the similarity measures described in the previous section in order to maximize the dissimilarity between the single-point mutants of the small RNA segments and their wildtype. Optimizing for dissimilarity is achieved by generating tables such as Table I for the sequence 5′-GAAAAUAAAAU-CUCAAAUGGUCAAUAUAUAUCGUAA GAUC-3′ of Fig. 4. We are aiming for the largest distance measures between the triggered secondary structure and the wildtype secondary structure relative to such distances in the other 102 sequences. For example, let us examine the sequence in Table I. Using EV2, it is clear that the mutation C14U may produce a conformational change from the predictive standpoint since in most other mutations the second eigenvalue of the Laplacian matrix remains 2.0 (a star shape with two vertices). Furthermore, in this example one also notices the benefit of using a nonmetric measure such as EV2, since from the mathematical theorems about EV2 in the case of tree-graphs we can immediately deduce without a graphical picture that the wildtype possesses a tree-graph shape with two vertices (the only combination in which EV2 may possibly become 2.0). Such information about the overall configuration of the RNA secondary structure may not be possible to obtain by examining numbers that convey distances in a metric measure. Likewise, using RNAdist, we again observe a jump from 0 to the highest value of 34 for the mutant C14U although this calculation that utilizes a fine-grain representation is independent from EV2 that uses a coarse-grain tree graph representation. We notice the same effect in the Wiener and HCT measures for C14U: relative to the other single mutations, a significantly large Wiener number of 4081.87 is reported with respect to the wildtype and a significantly small HCT value of 0.40 is reported relative to the wildtype. Indeed, inserting the mutation C14U to Vienna's *RNAfold*, we can observe that energy minimization predicts a conformational rearrangement (see Fig. 4). Thus, most times the various distance measures that we use are in agreement with each other. When one of the distance measures is not in agreement with the other three because of an exceptional case that may report a false positive corresponding to this measure, it is possible to verify the particular case using a direct simulation with *mfold* or *RNAfold*.

TABLE I
Analysis of the Small Switch Building Block Candidate Depicted in Fig. 4, Using Four Different Similarity Measures

Sequence Tag	EV2	Wiener	RNAdist	HCT
Small 1-WT	2	9384.69	0	1.00
Small 1-C14U	1	4081.87	34	0.40
Small 1-A7U	0.585786	5242.86	20	0.55

The resulting three synthetic riboswitches are depicted in Figs. 5–7. In each, the expression platform contains a small switch building block that responds to a single-point mutation by a conformational change when detached from the G-box domain that was chosen in advance.

VI. Conclusions

Motivated by the modularity of known natural riboswitches that consist of a ligand-binding sensor part attached to a small RNA switch component, we have described a computational procedure for designing a synthetic riboswitch with certain favorable properties. Starting from a pool of randomly generated

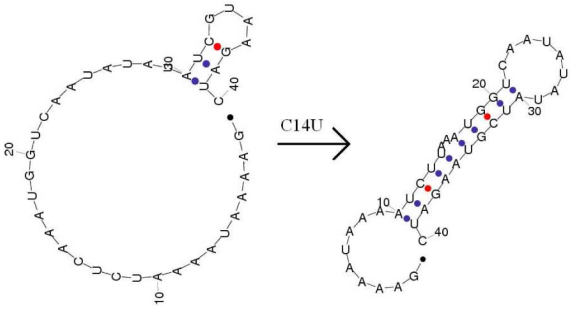


Fig. 4. A small switch building block that is predicted to respond to a single-point mutation by a conformational rearrangement. This particular building block will be inserted in the synthetic riboswitch Candidate 1.

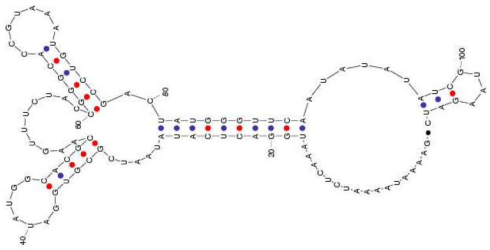


Fig. 5. Candidate 1: a possible computer-generated synthetic ribowitch (containing a G-box domain) with favorable mutational stability properties. The G-box starts in location 21 and ends in location 87. The resulting hairpin is a GUAA tetraloop.

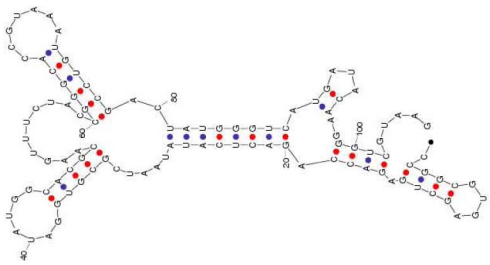


Fig. 6. Candidate 2: a possible computer-generated synthetic ribowitch (containing a G-box domain) with favorable mutational stability properties. The G-box starts in location 21 and ends in location 87. The resulting hairpin is a GUGA tetraloop.

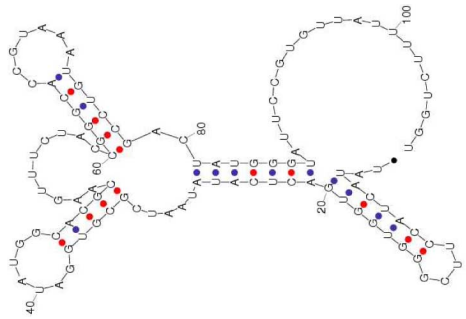


Fig. 7. Candidate 3: a possible computer-generated synthetic ribowitch (containing a G-box domain) with favorable mutational stability properties. The G-box starts in location 21 and ends in location 87. The resulting hairpin is a UUCG tetraloop.

RNA sequences, we have gradually screened our initial pool based on considerations of modularity, mutational robustness, and optimized switching capabilities of the small RNA building block that is situated in the expression platform. The prediction of switching capability for the building block candidates we have reached can be validated in laboratory experiments using the same procedures that were described for the RNA temperature building blocks. The current experimental results led us to opt for working with single-point mutation building block switches instead of temperature building block switches for the design of synthetic riboswitches.

The computational procedure is based on various pattern representations of predicted RNA secondary structures emanating from energy minimization methods, and similarity measures between these patterns as outlined in the computational methods section. Thus, throughout our calculations we rely on energy minimization predictions and to ascertain their correctness we work with reduced sequence lengths and generate building blocks that can be verified in "wet" laboratory experiments. Our preprocessing strategy, starting from random sequences and performing a computational screening to obtain synthetic riboswitch candidates, can be supplemented by further important considerations such as thermodynamic stability and predesigned aptamers. In our reported simulations we selected a particular G-box domain as our aptamer for a matter of convenience, but in general there should be no restriction on the aptamer selection or sequence lengths.

The continual discoveries of RNA switches and RNA natural sensors open a new era in nanobiotechnological aspects, where by an RNA sequence tailored for control and detection purposes or gene regulation can be assembled from building block components. Computer-aided design can be utilized to construct synthetic riboswitches with favorable properties and a variety of optimization considerations. Aside of aptamer design for the ligand-binding part, the rugged energy landscape of small RNA sequences can be used to select optimized building block components for the expression platform by taking into account thermostability and mutational robustness considerations in order to generate small RNA switches.

REFERENCES

[1] A. Avihoo and D. Barash, "Shape similarity measures for the design of small RNA switches," *J. Biomol. Struct. Dyn.*, vol. 24, no. 1, pp. 17–24, 2006.

[2] D. Barash, "Deleterious mutation prediction in the secondary structure of RNAs," *Nucl. Acids Res.*, vol. 31, no. 22, pp. 6578–6584, 2003.

[3] ——, "Second eigenvalue of the Laplacian matrix for predicting RNA conformational switch by mutation," *Bioinformatics*, vol. 20, no. 12, pp. 1861–1869, 2004.

[4] R. T. Batey, S. D. Gilbert, and R. K. Montange, "Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine," *Nature*, vol. 432, no. 7015, pp. 411–415, 2004.

[5] S. Bonhoeffer, J. S. McCaskill, P. F. Stadler, and P. Schuster, "RNA multi-structure landscape. A study based on temperature dependent partition functions," *Eur. Biophys. J*, vol. 22, pp. 13–24, 1993.

[6] R. R. Breaker, "Natural and engineered nucleic acids as tools to explore biology," *Nature*, vol. 432, pp. 838–845, 2005.

[7] P. Brion and E. Westhof, "Hierarchy and dynamics of RNA folding," *Annu. Rev. Biophys. Biomol. Struct*, vol. 26, pp. 113–137, 1997.

[8] B. Cohen and S. Skiena, "Natural selection and algorithmic design of mRNA," *J. Comput. Biol.*, vol. 10, pp. 419–432, 2003.

[9] M. Fiedler, "Algebraic connectivity of graphs," *Czechoslovak Math. J.*, vol. 23, pp. 298–305, 1973.

[10] C. Flamm, W. Fontana, I. L. Hofacker, and P. Schuster, "RNA folding at elementary step resolution," *RNA*, vol. 6, no. 3, pp. 325–338, 2000.

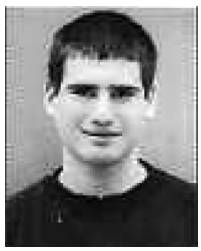
[11] C. Flamm, I. L. Hofacker, S. Maurer-Stroh, P. F. Stadler, and M. Zehl, "Design of multistable RNA molecules," *RNA*, vol. 7, pp. 254–265, 2001.

[12] R. Grone and R. Merris, "Algebraic connectivity of trees," *Czechoslovak Math. J.*, vol. 37, pp. 660–670, 1987.

[13] I. L. Hofacker, W. Fontana, P. F. Stadler, L. S. Bonhoeffer, M. Tacker, and P. Schuster, "Fast folding and comparison of RNA secondary structures," *Monatsh.Chem.*, vol. 125, pp. 167–188, 1994.

[14] I. L. Hofacker, "Vienna RNA secondary structure server," *Nucl. Acids Res.*, vol. 31, pp. 3429–3431, 2003.

[15] S. R. Holbrook, "RNA structure: The long and the short of it," *Curr. Opin. Struct. Biol.*, vol. 15, no. 3, pp. 302–308, 2005.
[16] K. A. LeCuyer and D. M. Crothers, "Kinetics of an RNA molecular switch," *Proc. Nat. Acad. Sci.*, vol. 91, pp. 3373–3377, 1994.
[17] A. Lescoute and E. Westhof, "Riboswitch structures: Purine ligands replace tertiary contacts," *Chem. Biol.*, vol. 12, no. 1, pp. 10–13, 2005.
[18] M. Mandal and R. R. Breaker, "Gene regulation by riboswitches," *Nat. Rev. Mol. Cell. Biol.*, vol. 5, no. 6, pp. 451–463, 2004.
[19] M. Mandal, B. Boese, J. E. Barrick, W. C. Winkler, and R. R. Breaker, "Riboswitches control fundamental biochemical pathways in bacillus subtilis and other bacteria," *Cell*, vol. 113, no. 5, pp. 577–586, 2003.
[20] H. Margalit, B. A. Shapiro, A. B. Oppenheim, and J. V. Maizel, "Detection of common motifs in RNA secondary structure," *Nucl. Acids Res.*, vol. 17, no. 12, pp. 4829–4845, 1989.
[21] D. H. Mathews, J. Sabina, M. Zuker, and D. H. Turner, "Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure," *J. Mol. Biol.*, vol. 288, pp. 911–940, 1999.
[22] J. S. McCaskill, "The equilibrium partition function and base pair binding probabilities for RNA secondary structure," *Biopolymers*, vol. 29, pp. 1109–1119, 1990.
[23] R. Merris, "Characteristic vertices of trees," *Lin. Multi. Alg.*, vol. 22, pp. 115–131, 1987.
[24] ——, "An edge version of the matrix-tree theorem and the Wiener index," *Lin. Multi. Alg.*, vol. 25, pp. 291–296, 1989.
[25] R. Micura and C. Höbartner, "On secondary structure rearrangements and equilibria of small RNAs," *ChemBioChem*, vol. 4, pp. 984–990, 2003.
[26] E. Nudler and A. S. Mironov, "The riboswitch control of bacterial metabolism," *Trends Biochem. Sci.*, vol. 29, pp. 11–17, 2004.
[27] J. H. Nagel and C. W. Pleij, "Self-induced structural switches in RNA," *Biochimie*, vol. 84, no. 9, pp. 913–923, 2002.
[28] J. H. Nagel, C. Flamm, I. L. Hofacker, K. Franke, M. H. de Smit, P. Schuster, and C. W. Pleij, "Structural parameters affecting the kinetics of RNA hairpin formation," *Nucl. Acids Res.*, vol. 34, no. 12, pp. 3568–3576, 2006.
[29] R. Penchovsky and R. R. Breaker, "Computational design and experimental validation of oligonucleotide-sensing allosteric ribozymes," *Nat. Biotechnol.*, vol. 23, pp. 1424–1433, 2005.
[30] D. Sankoff, J. B. Kruskal, S. Mainville, and R. J. Cedergren, "Fast algorithms to determine RNA secondary structures containing multiple Loops," in *Time Warps, String Edits, and Macromolecules: The Theory and Practice of Sequence Comparison*, D. Sankoff and J. B. Kruskal, Eds. Reading, MA: Addison-Wesley, 1983, pp. 913–923.
[31] P. Schuster, W. Fontana, P. F. Stadler, and I. L. Hofacker, "From sequences to shapes and back: A case study in RNA secondary structures," *Proc. R. Soc. Lond. B. Biol. Sci*, vol. 255, no. 1344, pp. 279–284, 1994.
[32] A. Serganov, Y. R. Yuan, O. Pikovskaya, A. Polonskaia, L. Malinina, A. T. Phan, C. Hobartner, R. Micura, R. R. Breaker, and D. J. Patel, "Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs," *Chem. Biol.*, vol. 11, no. 12, pp. 1729–1741, 2004.
[33] B. A. Shapiro, "An algorithm for comparing multiple RNA secondary structures," *Comput. Appl. Biosci.*, vol. 4, pp. 387–393, 1988.
[34] B. A. Shapiro and K. Z. Zhang, "Comparing multiple RNA secondary structures using tree comparisons," *Comput. Appl. Biosci.*, vol. 6, pp. 309–318, 1990.
[35] G. A. Soukup and R. R. Breaker, "Nucleic acid molecular switches," *Trends. Biotechnol.*, vol. 17, no. 12, pp. 469–476, 1999.
[36] N. Sudarsan, J. K. Wickiser, S. Nakamura, M. S. Ebert, and R. R. Breaker, "An mRNA structure in bacteria that controls gene expression by binding lysine," *Genes Develop.*, vol. 17, pp. 2688–2697, 2003.
[37] I. Tinoco and C. Bustamante, "How RNA folds," *J. Mol. Biol.*, vol. 293, no. 2, pp. 271–281, 1999.
[38] A. G. Vitreschak, D. A. Rodionov, A. A. Mironov, and M. S. Gelfand, "Riboswitches: the oldest mechanism for the regulation of gene expression?," *Trends Genet.*, vol. 20, no. 1, pp. 44–50, 2004.
[39] B. Voss, C. Meyer, and R. Giegerich, "Evaluating the predictability of conformational switching in RNA," *Bioinformatics*, vol. 20, no. 10, pp. 1573–1582, 2004.
[40] A. E. Walter, D. H. Turner, J. Kim, M. H. Lyttle, P. Muller, D. H. Mathews, and M. Zuker, "Coaxial stacking of helixes enhances bBinding of oligoribonucleotides and improves predictions of RNA folding," *Proc. Nat. Acad. Sci.*, vol. 91, no. 20, pp. 9218–9222, 1994, USA.
[41] H. Wiener, "Structural determination of paraffin boiling points," *J. Amer. Chem. Soc.*, vol. 69, pp. 17–20, 1947.
[42] W. C. Winkler, S. Cohen-Chalamish, and R. R. Breaker, "An mRNA structure that controls gene expression by binding FMN," *Proc. Nat. Acad. Sci. USA.*, vol. 99, no. 25, pp. 15 908–15 913, 2002.
[43] M. T. Wolfinger, W. A. Svrcek-Seiler, C. Flamm, I. L. Hofacker, and P. F. Stadler, "Efficient computation of RNA folding dynamics," *J. Phys. A*, vol. 37, no. 17, pp. 4731–4741, 2004.
[44] M. Zuker, "On finding all suboptimal foldings of an RNA molecule," *Science*, vol. 244, pp. 48–52, 1989.
[45] ——, "Calculating nucleic acid secondary structure," *Curr. Opin. Struct. Biol.*, vol. 10, no. 3, pp. 303–310, 2000.
[46] ——, "Mfold web server for nucleic acid folding and hybridization prediction," *Nucl. Acids Res.*, vol. 31, pp. 3406–3415, 2003.



Assaf Avihoo received the B.Sc. degree from the Department of Computer Science at Ben Gurion University of the Negev, Beer-Sheva, Israel, in 2003.

Since September 2003, he has been a Graduate Student in the Department of Computer Science, Ben Gurion University of the Negev. He is teaching in the systems programming laboratory, and his research interests include bioinformatics, RNA structure predictions, and scientific computing.

Idan Gabdank received the B.Sc. degree from the Department of Computer Science at Ben Gurion University of the Negev, Beer-Sheva, Israel, in 2004.

Since September 2004, he has been a Graduate Student in the Department of Computer Science at Ben Gurion University of the Negev. His research interests include bioinformatics, molecular evolution, RNA structure predictions, RNA structure determination by biochemical probing, and scientific computing.

Michal Shapira received the Ph.D. degree in life sciences in 1984 from the Weizmann Institute of Science, Rehovoth, Israel.

Between 1984 and 1986 she was a Postdoctoral Fellow at the University of Michigan, Ann Arbor, working on regulation of herpes virus genes. From 1986 to 1995 she worked at the Weizmann Institute of Science as an Assistant Professor, on molecular genetics of the protozoan parasite Leishmania, with a special focus on mechanisms of developmental gene regulation in this organism. During 1995 she was a Visiting Professor at Duke University, Durham, NC, and since 1995 she is at the Ben Gurion University of the Negev, Beer Sheva, Israel, currently an Associate Professor, working on regulatory mechanisms of stress- induced gene expression in microorganisms.

Danny Barash (M'01) received the Ph.D. degree in applied science in 1999 from the University of California, 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 of the Negev, Beer-Sheva, Israel, 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.