Predicting Folding Pathways between RNA Conformational Structures Guided by RNA Stacks

Yuan Li
Department of Electrical Engineering and Computer Science
University of Central Florida
Orlando, FL, USA, 32816-2362
liy@eecs.ucf.edu

Shaojie Zhang
Department of Electrical Engineering and Computer Science
University of Central Florida
Orlando, FL, USA, 32816-2362
shzhang@eecs.ucf.edu

ABSTRACT

In this paper, we propose a new algorithm for predicting low energy barrier folding pathways between conformational secondary structures of a single RNA molecule. Most existing heuristic algorithms guide the construction of folding pathways by free energies of intermediate structures in the next move during the folding. However due to the size and ruggedness of RNA energy landscape, energy-guided search can become trapped in local optima. Here, we propose an algorithm that guides the construction of folding pathways through the formation and destruction of RNA stacks. Guiding the construction of folding pathways by coarse grained movements of RNA stacks can help reduce the search space and make it easier to jump out of local optima. RNAEAPath is able to find lower energy barrier folding pathways between secondary structures of conformational switches and outperforms the existing heuristic algorithms in most test cases. The source code of RNAEAPath and associated supplementary data are available at http://genome.ucf.edu/RNAEAPath.

Categories and Subject Descriptors

H.4 [**LIFE AND MEDICAL SCIENCES**]: Biology and Genetics; I.2.8 [**ARTIFICIAL INTELLIGENCE**]: Problem Solving, Control Methods, and Search—*Heuristic methods*

General Terms

Algorithm, Theory

Keywords

Conformational secondary structure, RNA folding pathway, RNA energy barrier, evolutionary algorithm

1. INTRODUCTION

RNA molecules play critical roles in the cell. The secondary structures of RNA molecules have been extensively studied because they provide insights into the functionality of RNAs. Native (functional) RNA secondary structures are usually thermodynamically stable and many of them are also the minimum free energy (MFE) structures. Nevertheless, at times, RNA molecules may fold into alternative secondary structures in order to participate in certain biological processes. For example, the SV-11 RNA folds into a metastable conformational structure and acts as a template for its own replication using Qβ replicase [3, 4]. Further, RNA conformational switches can transform between alternative secondary structures dynamically in response to various environmental stimuli (such as heat shock and cold shock) [6, 16, 24, 25], and carry out RNA-mediated biological activities, such as switching on or off downstream gene translation activities [22, 29, 37], regulating RNA splicing via multiple-state splicesomal conformations [32], and regulating the life cycles of virus [31].

The conformational transformations between alternative structures involve the folding of an RNA molecule into a series of sequential adjacent intermediate structures [18]. RNA folding pathways provide valuable information for understanding the catalytic and regulatory functions of RNAs (such as hok/sok of plasmid R1 [14]). RNA folding pathways may also impact subsequence biological events (such as formation of tertiary structures). Furthermore, prediction algorithms can help the design of RNA switches by providing prescribed structural alternatives.

In this paper, we present a new approach, RNAEAPath, for computing near optimal direct or indirect folding pathways between two secondary structures of an RNA molecule. We guide the search for low energy barrier folding pathways by integrating a variety of strategies for simulating the formation and destruction of RNA stacks in a flexible framework. Benchmark tests on conformational switches show that RNAEAPath produces lower energy barrier folding pathways and outperforms the existing heuristic approaches in most test cases.

1.1 Preliminary

Consider an RNA sequence as a string $x = x_1 \cdots x_n$ of n letters over alphabet $\Sigma = \{\mathrm{A}, \mathrm{U}, \mathrm{G}, \mathrm{C}\}$. A pair of complementary nucleotides x_i and x_j, can form hydrogen bonds and interact with each other, denoted by $x_i \cdot x_j$. In this paper, we only consider the canonical base pairings ($A \cdot U$ and G $\cdot$ C) and the wobble base pairing (G $\cdot$ U). A *secondary structure* S of the RNA sequence x is a set of disjoint paired bases (i, j), where $1 \leq i < j \leq n$. S may be represented by a length n string of dots and brackets, where dots represent unpaired bases and brackets represent paired bases. An RNA structure can comprise of *stacks* which are lists of consecutive base pairs ($\{(i, j), (i + 1, j - 1), \ldots, (i + w, j - w)\}$ such that $x_i \cdot x_j, \ldots, x_{i+w} \cdot x_{j-w}$), and unstacking base pairs. A secondary structure is *pseudoknotted* if it contains two base pairs (i, j) and (i', j') with $i < i' < j < j'$. In this paper, we only con-

ACM BCB '11 Chicago Illinois

sider pseudoknot-free structures. A base pair is compatible with a secondary structure if the base pair can be added to the structure without leading to a pseudoknotted structure or pairing a base with more than one partner. A stack is compatible with S if each base pair in the stack is either in S or is compatible with S.

The free energy of a secondary structure S is denoted by $E(S)$. The set of *neighboring structures* of S consists of all structures that differ from S by an addition or deletion of exactly one base pair. For two secondary structures A and B, the *distance* between A and B is the number of base pairs in A not in B plus the number of base pairs in B not in A (i.e. $|(A-B) \cup (B-A)|$). A *folding pathway* from A to B is a sequence of intermediate structures $A = S_0, \ldots, S_m = B$ such that for all $0 \leq i < m$, intermediate structure S_{i+1} is a neighboring structure of S_i. A folding pathway is *direct* if the intermediate structures contain only base pairs in A and B (i.e. $S_i \subseteq A \cup B$ for $1 \leq i < m$) and otherwise is *indirect*. The *saddle point* of a pathway is an intermediate structure with the highest energy, and the *energy barrier* of a pathway is the energy difference between its saddle point and the initial structure. Since the folding of RNA structures is thermodynamically-driven and tends to avoid high-energy intermediate structures, current computational methods aim to find RNA folding pathways with the lowest energy barriers.

1.2 Previous studies

A lot of research has been done on predicting low energy barrier folding pathways. Morgan and Higgs proposed a greedy algorithm that employs the Nussinov model [27, 28] for computing direct folding pathways with minimum energy barrier. They also described a heuristic that samples low energy structures from the partition function and glues them together by direct pathways [23]. The Nussinov model is simple and easy to implement, in which base stacking and loop entropies have no energetic contributions. Based on this model, Thachuk *et al.* [34] developed an exact algorithm, PathwayHunter, which exploits elegant properties of bipartite graphs for finding the globally optimal direct pathways. However, the Nussinov model is not as accurate as the Turner energy model [21, 35] for approximating RNA thermodynamics. An exact solution based on the Turner energy model is also available. BARRIERS [9, 11], exactly computes the globally optimal folding pathways between any two locally optimal secondary structures. BARRIERS reads an energy sorted list of RNA secondary structural conformations produced by RNAsubopt [38] and is able to compute both direct and indirect low energy barrier pathways.

Nevertheless, the above exact solutions are all exponential in time, because the problem itself is NP-hard [20]. Many heuristic algorithms have also been proposed following the seminal work of Morgan and Higgs. Flamm *et al.* [10] used breadth-first search in their heuristics (in Vienna RNA Package [15]) and kept the best k candidates at each step to bound the search. Voss *et al.* [36] devised a straightforward strategy for greedily searching direct pathways. Geis *et al.* [13] described a greedy heuristic to explore the search space of direct pathways and they also integrated look ahead techniques to diminish the search space. Recently, Dotu *et al.* [7] developed RNATabuPath, a fast heuristic that employs a TABU semi-greedy search to construct near optimal (both direct and indirect) folding trajectories. In addition, other heuristic approaches, by splitting the pathways into shorter pathways and solving each individually, have also been proposed [5, 17].

Many of the existing heuristic algorithms start from an initial structure A, and, at each single step i, walk from the intermediate structure S_i to one of its neighbors S_{i+1} until finally the end structure B is reached. The definition of neighborhood relationships as well as the fitness functions can be different. The *fitness function* of S_i is usually defined on the free energy of S_i, or the distance from S_i to B, or a function of both. In general, greedy algorithms select the 'best' neighbor structure that has the best fitness. In contrast, semi-greedy algorithms may select any one from the top k structures for randomization. RNATabuPath, which is more sophisticated and outperforms other methods [7], keeps a tabu list for saving recently taken moves such that they can not be applied in certain steps until being removed from the tabu list. In general, during the construction of a folding pathway, these heuristic algorithms select the next intermediate structures from a set of neighboring structures that have the top lowest free energy or have the top shortest distance to B (or the combination of both).

1.3 Motivations

However, using energy to guide the construction of folding pathways in the above-mentioned heuristic algorithms has its downsides. The RNA energy landscapes can be extremely large and rugged [30, 31] and the ruggedness of RNA energy landscape may cause the energy-guided search to become trapped in a local optimum. Similar to using structural rearrangements for modeling RNA folding kinetics [26], we want to construct candidate folding pathways in a manner that make it easier to jump out of local optima. It has been revealed that stacking base pairs contribute significantly to the stabilization of RNA secondary structures [33, 39]. The dominant RNA folding pathways involve the formation and destruction of the stacks, and the cooperative formation of a stack along with the partial melting of an incompatible stack [40]. In this paper, we propose to guide the construction of pathways by the formation and destruction of stacks (not by free energy or by distance to the end structure). We still select the constructed folding pathways according to their energy barriers. Although the construction of folding pathways is not driven by thermodynamics, the selection of folding pathways is based on energy barriers. Guiding the construction of folding pathways by coarse grained movements of RNA stacks may help reduce the search space and makes it easier to jump out of local optima.

In the rest of this paper, Section 2 describes the representation of folding pathways and the detailed strategies employed by RNAEAPath. Section 3 presents benchmarking results of RNAEAPath against existing methods followed by concluding remarks in Section 4.

2. METHOD

2.1 Representation of RNA folding pathways

Given an initial structure A and an end structure B, we use a sequence of *actions* successively applied to A, rather than a sequence of intermediate structures, to represent a folding pathway from A to B. Representing a pathway by an action chain can avoid cyclic additions and deletions of base pairs and make it easy to simulate the formation and deletion of RNA stacks. A similar representation has also been employed in the previous work of Thachuk *et al.* [34].

We use two types of actions, $\texttt{add}_{i,j}$ and $\texttt{del}_{i,j}$ in the representation of RNA folding pathways. For an intermediate secondary structure S of an RNA sequence x, the action $\texttt{add}_{i,j}$ denotes the 'add'ition of base pair (i,j) to S (i.e. $\texttt{add}_{i,j}(S) = S \cup \{(i,j)\}$) and $\texttt{del}_{i,j}$ denotes the 'del'etion of base pair (i,j) from S (i.e. $\texttt{del}_{i,j}(S) = S - \{(i,j)\}$). An action is *direct* if it concerns a base pair in $A \cup B$ and *indirect* otherwise. The simplest direct pathways from A to B concern sequential deletions of all base pairs in $A-B$ followed by additions of all base pairs in $B-A$.

Consider an example sequence $x =$ GGGGAAAACCCCUUUU

	Structures GGGGAAAACCCCUUUU	Energy (kcal/mol)	Actions	
A	((((....))))....	-6.60	a_1	$\texttt{del}_{1,12}$
S_1	.(((....))).....	-2.90	a_2	$\texttt{del}_{2,11}$
S_2	..((....)).....	0.40	a_3	$\texttt{del}_{3,10}$
S_3	...(....).....	3.70	a_4	$\texttt{del}_{4,9}$
S_4	....	0.00	a_5	$\texttt{add}_{8,13}$
S_5	.......(....)...	5.50	a_6	$\texttt{add}_{7,14}$
S_6	......((....))..	4.60	a_7	$\texttt{add}_{6,15}$
S_7	.....(((....))).	3.70	a_8	$\texttt{add}_{5,16}$
B	....((((....))))	2.80		

Figure 1: A simple folding pathway that converts an RNA sequence from structure A to B. The leftmost column shows a simple direct pathway from A to B, the center column shows the free energies (in kcal/mol) of the intermediate structures, and the rightmost column presents the action chain $a_1, \ldots, a_8$ for this pathway.

with initial and final structures shown in Figure 1. This simple pathway is obtained by first deleting all GC pairs from A until the RNA is single stranded, and then adding all AU pairs until B is obtained. Note that each intermediate structure S_i differs from both its successor and predecessor by exactly one base pair. The actions in the example are all direct actions and the energy barrier is $5.50 - (-6.60) = 12.10$ kcal/mol.

An addition action $\texttt{add}_{i,j}(S)$ *conflicts with* S if either x_i or x_j is already paired in S, and it *clashes with* S if there exists a base pair $\{(x'_i, x'_j) \in S | i < i' < j < j'$ or $i' < i < j' < j\}$. A deletion action $\texttt{del}_{i,j}(S)$ *conflicts with* S if $(x_i, x_j) \notin S$. An addition or deletion action is *valid* and can be applied to S *properly* if it neither conflicts with nor clashes with S.

A pathway from A to B can be represented by an *action chain*, which is a sequence of valid actions $a_1, \ldots, a_m$ such that $S_0 = A$, $S_t = a_t(S_{t-1})$ for $1 \leq t \leq m$ and $S_m = B$. Note that an action chain for A to B implies a sequence of valid actions that can be successively applied to A without introducing conflicts or clashes and produce B. We use the term "action chain" when the sequence is certified to be valid, and the term "sequence of actions" if its validity is not guaranteed.

This representation of a pathway p from A to B has the following important properties. First, every folding pathway can be represented by a unique action chain and every action chain represents a unique folding pathway (note that it is not necessarily true for a sequence of actions). Second, rearranging the order of actions in p results in a new sequence of actions which represents a new folding pathway from A to B when it is *valid*. (It is an action chain that can be successively applied to A properly and obtain B.) Third, introducing a pair of complementary actions (e.g. $\texttt{add}_{i,j}$ and $\texttt{del}_{i,j}$) to p results in a new sequence of actions which also represents a new folding pathway from A to B if it is *valid*.

In RNAEAPath, folding pathways are represented in the form of action chains, instead of a sequence of intermediate structures. This representation makes the life cycle of a folding pathway transparent to the algorithm and also makes it easier for us to simulate the cooperative formation and destruction of RNA stacks by re-arranging the order of actions or introducing multiple pairs of complementary actions.

2.2 Predicting low energy barrier folding pathways

Given an RNA sequence x, an initial structure A and a final structure B, RNAEAPath computes a near optimal low energy barrier folding pathway from A to B in an evolutionary algorithm framework [8]. Figure 2 elucidates the overall paradigm for RNAEAPath. In this algorithm, the population of each generation is comprised of folding pathways ordered by their *fitness* which will be defined in section 2.2.1. The functions $\mathbb{M}_y(p)$ are *mutation strategies*, each of which takes in a pathway p and produces a set of offspring pathways. These mutation strategies are central to the effectiveness of RNAEAPath and will be discussed in section 2.3. ℓ_1, ℓ_2, ℓ_3, MAX and γ are positive integer control parameters, and the default values are shown in section 3.2.

Procedure: RNAEAPath(x, A, B)
1: $\Delta \leftarrow |E(B) - E(A)|$
2: $k \leftarrow 0$
3: Initialize $\mathbb{P}_0$ and sort individuals in it by energy barriers
4: $\text{OPT}_0 \leftarrow \mathbb{P}_0[1]$
5: **while** !STOP(k, OPT, Δ) **do**
6: $\quad k \leftarrow k + 1$
7: $\quad \mathbb{O}_k \leftarrow \mathbb{P}_{k-1}[1 \ldots \ell_1]$
8: $\quad$ **for all** $p \in \mathbb{P}_{k-1}$ **do**
9: $\quad\quad \mathbb{T} \leftarrow \left(\bigcup_{y=1}^{Y} \mathbb{M}_y(p)\right)$
10: $\quad\quad \mathbb{O}_k \leftarrow \mathbb{O}_k \cup \mathbb{T}[1 \ldots \ell_2]$
11: $\quad$ **end for**
12: $\quad \text{OPT}_k = \mathbb{O}_k[1]$
13: $\quad \mathbb{P}_k \leftarrow \mathbb{O}_k[1 \ldots \ell_3]$
14: **end while**
15: **return** OPT_k

Figure 2: Overview of RNAEAPath. In this procedure, the input is an RNA sequence x with the start and end structures A and B, and the output is the best folding pathway in k iterations (OPT_k). For notations, k is the number of iterations, $\mathbb{P}_0$ is the initial population, and $\mathbb{P}_k$ is the folding pathway population of the k^{th} iteration. $\mathbb{T}$ contains all the offspring folding pathways produced by applying mutation strategies $\mathbb{M}_1, \ldots, \mathbb{M}_Y$ to each pathway p in the $(k-1)^{st}$ population. $\mathbb{O}_k$ is an ordered list of offspring folding pathways of the k^{th} generation, from which, the population for the next iteration ($\mathbb{P}_{k+1}$) is selected. Folding pathways in $\mathbb{P}_k$ and $\mathbb{O}_k$ are sorted based on their fitness and $\mathbb{P}_k[1 \ldots \ell]$ are the top ℓ best folding pathways in $\mathbb{P}_k$.

The initial population of RNAEAPath, $\mathbb{P}_0$, is filled with a set of simple pathways described in Section 2.2.2. Then, the algorithm goes through several iterations. $\mathbb{P}_{k-1}$ is the population of the $k - 1^{st}$ iteration. In the k^{th} iteration, the algorithm produces $\mathbb{O}_k$ (an ordered list of pathways) and $\mathbb{P}_k$ (the population of the k^{th} iteration) from $\mathbb{P}_{k-1}$. $\mathbb{O}_k$ stores the best ℓ_1 pathways in $\mathbb{P}_{k-1}$ and the best ℓ_2 pathways produced by each $p \in \mathbb{P}_{k-1}$. More specifically, each pathway $p \in \mathbb{P}_{k-1}$ produces t_y^k offsprings through every mutation strategy $\mathbb{M}_y$ ($1 \leq y \leq Y$). The resulting offsprings produced by p are stored in a temporary list $\mathbb{T}$, and the top ℓ_2 pathways are added to $\mathbb{O}_k$. Finally, the best solution of the k^{th} iteration, termed as OPT_k, is the best pathway in $\mathbb{O}_k$. And, $\mathbb{P}_k$ (the population of the k^{th} iteration) is composed of the best ℓ_3 pathways of $\mathbb{O}_k$ and will be used in the next iteration to produce $\mathbb{P}_{k+1}$. This helps keep the diversity of the population large, since $\mathbb{P}_k$ contains at most ℓ_2 offsprings produced by each $p \in \mathbb{P}_{k-1}$, no matter how many high-qualified offsprings are produced by each pathway. The algorithm terminates when a stopping condition is met, and it returns the best solution of the last iteration. Since $\mathbb{O}_k$ retains the best ℓ_1 pathways from $\mathbb{P}_{k-1}$ in each iteration, the best one ever encountered by the algorithm is retained in lists $\mathbb{O}_k$ and $\mathbb{P}_k$, and

Structures GGGGGGAAAAACCCCCC	E(S) (kcal/mol)	Structures GGGGGGAAAAACCCCCC	E(S) (kcal/mol)
.....	0	.....	0
.....(.....).....	3.7	(.....)	4.04
....((.....)).....	0.4	(.(.....).)	4.10
...(((.....)))...	-2.9	(.(.(.....).).)	3.8
.(((((.....))))).	-9.5	(((.((.....)).)))	-5.0
((((((.....))))))	-12.0	((((((.....))))))	-12.0

Figure 3: Two different folding pathways that form an identical stack. Left: The stack is formed successively. Right: The stack is constructed by random formation of base pairs. The right pathway yields a higher energy barrier because the randomly introduced base pairs form unpaired loop regions that result in additional entropic penalties.

stored in OPT_k. So, OPT_k has no worse fitness when compared to OPT_{k-1}, and RNAEAPath always returns the best action chain it ever discovered.

In the remaining of this section, we discuss details regarding fitness evaluation, initialization of the population, stopping conditions and mutation strategies of RNAEAPath.

2.2.1 Fitness of action chains

The order of folding pathways (valid action chains) is primarily determined by their energy barriers. In case of a tie, the order is determined by the average of energy differences between the initial structure A and intermediate structures. Note that lower energies are preferred in the previous two methods of ordering. If a tie still exists, then shorter action chains are preferred. Action chains are ordered arbitrarily if their relative order can not be determined based on these three criteria.

2.2.2 The initial population of folding pathways

The initial population, $\mathbb{P}_0$, contains 4 *simple* pathways from A to B formed by first deleting all base pairs in $A - B$ and then adding those in $B - A$, similar to the pathway shown in Figure 1. Although we can also arrange base pair deletions and additions in an arbitrary order, we tailor them in a manner that simulates successive degradation and formation of RNA stacks. This is because random deletions and additions of base pairs tend to form additional unpaired loop regions that introduce entropic penalties (see Figure 3 for an illustration). We can degrade or form each stack either from the outmost base pair to the innermost base pair or vice verse. Usually, it yields a lower energy barrier if we degrade a stack from the outmost base pair to the innermost base pair and form a stack from the innermost base pair to the outmost base pair. However, for the sake of simplicity and generosity, we construct 4 simple pathways in $\mathbb{P}_0$, which degrade all the stacks from the same direction and form all the stacks from the same direction. These simple pathways constitute a diversified and unbiased initial population for the algorithm start from.

2.2.3 The number of offsprings produced by each mutation strategy

In each generation, the expected total number of offsprings produced by each individual is a constant positive integer $\mathfrak{L}$. The number of offsprings that each individual produces using mutation strategy $\mathbb{M}_y, (1 \leq y \leq Y)$, in the k^{th} generation, is denoted by ${\ell_{\mathbb{M}_y}}^k$. The value of ${\ell_{\mathbb{M}_y}}^k$ is determined adaptively according to the quality of the offsprings produced using $\mathbb{M}_y$ in the $k-1^{st}$ iteration. If $\mathbb{M}_y$ has a good performance in the previous generation, ${\ell_{\mathbb{M}_y}}^k$ grows accordingly; otherwise, it decreases. (See supplementary data for detailed descriptions.)

2.2.4 Stopping conditions

The algorithm terminates when (1) the current best solution achieves the lowest possible value $|E(B) - E(A)|$, or (2) when no improvement has been found over γ consecutive iterations (a plateau), or (3) when MAX number of iterations have passed and successive iterations do not discover better results. Note that the algorithm may simulate further than MAX iterations if improvements are made in the very last iteration and it stops immediately if no improvement is made between successive iterations. More specifically, the algorithm stops when any of the following conditions is satisfied:

1. the energy barrier of OPT_k is equivalent to $|E(B) - E(A)|$.
2. $k > \gamma$ and the fitness of OPT_k is equivalent to that of $\text{OPT}_{k-\gamma}$.
3. $k \geq MAX$ and the fitness of OPT_k is equivalent to that of OPT_{k-1}.

2.3 Mutation strategies

In RNAEAPath, the mutation strategies employed to evolve folding pathways can be categorized into three types: (1) rearranging the order of actions, (2) introducing indirect pathways and (3) formation of a single stack or cooperative conversion of a pair of incompatible stacks. In this section, let $\mathbb{M}_1, \ldots, \mathbb{M}_Y$ denote the mutation strategies and let $p = a_1, \ldots, a_m$ denote the input pathway $A = S_0, \ldots, S_m = B$. For each mutation strategy $\mathbb{M}_y(p)$, we describe the process for generating one new pathway q using each mutation strategy when given p.

2.3.1 Type 1: reordering of actions

As described in Section 2.1, shuffling the order of actions of the input pathway p can result in a new pathway from A to B. In RNAEAPath, two mutation strategies of this type are employed. $\mathbb{M}_1$ changes the position of an arbitrary action, and $\mathbb{M}_2$ swaps the positions of two arbitrary actions.

$\mathbb{M}_1$: Let $\mathbb{M}_1^{t_1,t_2}(p)$ denote the sequence of actions obtained by first removing an action a_{t_1} $(1 \leq t_1 \leq m)$ from p and then inserting it after a_{t_2}, for all $t_2 \in \{0, \ldots, t_1 - 1, t_1 + 1, \ldots, m\}$. Note that the resulting sequence of actions may not necessarily be a valid action chain. For instance, in Figure 1, $\mathbb{M}_1^{1,4}(p) = a_2, a_3, a_4, a_1, a_5, \ldots, a_8$ and $\mathbb{M}_1^{3,2}(p) = p$ are valid action chains, while $\mathbb{M}_1^{8,1}(p) = a_1, a_8, a_2, \ldots, a_7$ is not.

The procedure for computing $\mathbb{M}_1^{t_1,t_2}(p)$ is described in the following.

1. Choose t_1 uniformly at random from the interval $[1, m]$.
2. Compute the interval $[l, u]$, $(t_1 < l < u < m)$, where l is the minimum and u is the maximum such that for all $t_2 \in [l, u]$ and $t_2 \neq t_1$, $\mathbb{M}_1^{t_1,t_2}(p)$ is a valid action chain.
3. Choose t_2 from the interval $[l, u]$.
   - 3.1. If a_{t_1} is an addition operation, for all $l \leq t < t' \leq u$ and $t \neq t' \neq t_1$, the probability of choosing t is greater than that of t'.
   - 3.2. Otherwise (a deletion operation), for all $l \leq t < t' \leq u$ and $t \neq t' \neq t_1$, the probability of choosing t is less than that of t'

We do not choose t_2 $(t_2 \neq t_1)$ uniformly at random in $[l, u]$, instead, we tend to place addition operations in the front part of p, and deletion operations in the later part of p. This is because adding base pairs early and deleting them late during the folding may help stabilize the intermediate secondary structures. (Please see supplementary data for the detailed description of the discrete probability.)

$\mathbb{M}_2$: Let $\mathbb{M}_2^{t_1,t_2}(p)$ denote the sequence of actions obtained by swapping a_{t_1} with a_{t_2}. If the resulting sequence of actions is a valid action chain, let it be q; otherwise, restart the process. For example, in Figure 1, $\mathbb{M}_2^{1,8}(p)$ is not a valid action chain, while $\mathbb{M}_2^{2,4}(p) = a_1, a_4, a_3, a_2, a_5, \ldots, a_8$ is. t_1 and t_2 are chosen uniformly at random from $\{(t_1, t_2) : 1 \leq t_1 < t_2 \leq m\}$.

Mutation strategies of type 1 provide methods for shuffling the order of actions of an input pathway and generating slightly different new pathways. However, these strategies are not capable of introducing additional (indirect) base pairs, and the offsprings of a direct pathway produced through type 1 strategies are also direct. In the following, we will describe mutation strategies that are able to construct indirect pathways from a direct pathway.

2.3.2 *Type 2: introducing indirect pathways by adding a pair of complementary actions*

Morgan and Higgs [23] pointed out that the optimal folding paths are generally indirect pathways. This idea was further described by Dotu *et al.* [7]. The temporary formation of base pairs, especially those base pairs that do not belong to $A \cup B$, may lower the energies of intermediate structures and thus render better folding pathways. Similarly, temporary deletion and reformation of a base pair also can create an indirect pathway.

$\mathbb{M}_3$: Let $\mathbb{M}_3^{t_1,t_2,+(i,j)}(p)$ denote the sequence of actions obtained by introducing an addition action $\texttt{add}_{i,j}$ after a_{t_1} and its complementary action $\texttt{del}_{i,j}$ after a_{t_2}. Let $\mathbb{M}_3^{t_1,t_2,-(i,j)}(p)$ denote the sequence of actions obtained by introducing a deletion action $\texttt{del}_{i,j}$ after a_{t_1} and its complementary action $\texttt{add}_{i,j}$ after a_{t_2}. For example, in Figure 1, $\mathbb{M}_3^{1,7,+(1,16)}(p) = a_1, \texttt{add}_{1,16}, a_2, \ldots, a_7, \texttt{del}_{1,16}, a_8$. The procedures for computing $\mathbb{M}_3^{t_1,t_2,+(i,j)}(p)$ and $\mathbb{M}_3^{t_1,t_2,-(i,j)}(p)$ are similar to each other. In the following, we only describe the procedure for computing $\mathbb{M}_3^{t_1,t_2,+(i,j)}(p)$.

1. Choose t_1 uniformly at random from the interval $[1, m]$, and obtain the associated intermediate structure S_{t_1}.
2. Find a set of base pairs that neither conflict with nor clash with S_{t_1} and choose a base pair (i, j) uniformly at random from the set.
3. Compute the interval $[l, u]$, $(t_1 < l < u < m)$, where l is the minimum and u is the maximum such that for all values $t_2 \in [l, u]$ the resulting sequence of actions of $\mathbb{M}_3^{t_1,t_2,+(i,j)}(p)$ is a valid action chain.
4. Choose t_2 from the interval $[l, u]$ with the probability of choosing t greater than that of t' for all $t > t'$. (This is because (i, j) is not likely to be deleted soon after its formation.)

Mutation strategy $\mathbb{M}_3$ is capable of producing an indirect pathway from a direct pathway. In addition, a proper combination of multiple applications of $\mathbb{M}_3$ may result in a pathway which simulates the successive formation and deletion of a temporary stack during the folding. Take the pathway p in Figure 1 as an example, we can construct a pathway q that forms a temporary stack consisting of all the GU base pairs via a multiple application of $\mathbb{M}_3$, $q = \mathbb{M}_3^{5,7,+(3,14)}(\mathbb{M}_3^{3,7,+(2,15)}(\mathbb{M}_3^{1,7,+(1,16)}(p)))$.

2.3.3 *Type 3: formation of a single stack or simultaneous formation and deletion of a pair of incompatible stacks*

In this section, we will introduce mutation strategies for producing pathways that involve with formation and deletion of stacks. To perform this type of strategies, we first need to find all possible stacks in an RNA sequence x. We use the algorithm of Bafna *et al.* [2] to find the set of all possible stacks with more than 3 consecutive base pairs, and denote it by $STA(x)$. There are two strategies in Type 3: formation of a single stack ($\mathbb{M}_4$) and simultaneous formation and destruction of a pair of incompatible stacks ($\mathbb{M}_5$)

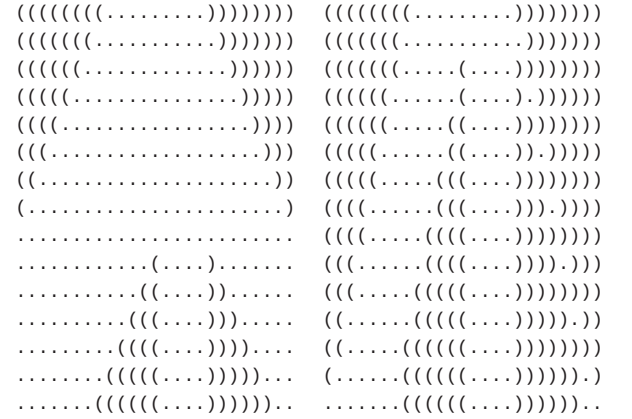

Figure 4: Two different folding pathways with the identical initial and final secondary structures. The left column: a simple folding pathway destroys a stack completely before an incompatible stack is formed. The right column: a folding pathway destructs a stack and forms an incompatible stack simultaneously.

$\mathbb{M}_4$: Let $\mathbb{M}_4^{t,h}(p)$ denote the sequence of actions obtained by forcing the formation of a stack $stack_h \in STA$ after action a_t, where $stack_h$ is compatible with S_t. The following describes the procedure for computing $\mathbb{M}_4^{t,h}(p)$.

1. Choose t uniformly at random from the interval $[1, m]$, and obtain the associated intermediate structure S_t.
2. Find a set of stacks that neither conflict with nor clash with S_t, and pick up a stack $stack_h$ uniformly at random from the set.
3. Ensure that each base pair (i, j) in $\{stack_h - S_t\}$ is sequentially (from the innermost base pair to the outmost base pair) formed after a_t.
   - 3.1. If an action $\texttt{add}_{i,j}$ appears in $\{a_{t+1}, \ldots, a_m\}$, move it up and place it after a_t using strategy $\mathbb{M}_1$.
   - 3.2. Otherwise, introduce a pair of complementary actions $\texttt{add}_{i,j}$ and $\texttt{del}_{i,j}$ to p after a_t using strategy $\mathbb{M}_3$.

We can introduce additional stacks that are compatible with S_t using $\mathbb{M}_4$ by forcing a sequence of addition actions successively forming base pairs in $\{stack_h - S_t\}$, after a_t. For example, the red stack in Figure 5 is constructed using $\mathbb{M}_4$.

$\mathbb{M}_5$: Let $\mathbb{M}_5^{t,h}(p)$ denote the sequence of actions obtained by forcing the formation of a stack $stack_h \in STA$ which is incompatible with S_t, after action a_t. Shown on the right side of Figure 4 is a folding pathway which simultaneously destructs and forms a pair of incompatible stacks. Shown on the left side is a simple folding pathway which has exactly the same start and end structures, while it folds into a single stranded structure during the folding. Usually, the pathway on the right has lower energy barrier than the one on the left because it never folds into a single stranded structure. The folding pathway on the right side of Figure 4 can be introduced using strategy $\mathbb{M}_5$. And, the procedure for computing $\mathbb{M}_5^{t,h}(p)$ is as follows:

1. Choose an arbitrary deletion action $a_t = \texttt{del}_{i,j}$ from p, and obtain the associated intermediate structure S_t.
2. Find a set of stacks which either conflicts with or clashes with S_t, and choose a stack $stack_h$ uniformly at random from the set.

Table 1: Benchmarks of BARRIERS, PathwayHunter, Findpath, RNATabuPath, and RNAEAPath for predicting folding pathways between conformational switches. Energy barriers (measured in kcal/mol) of the best folding pathways over n runs are shown. Boldface numbers are the best energy barriers found by the *heuristic* algorithms.

Instance	BARRIERS	PathwayHunter	FindPath	RNATabuPath	RNAEAPath	
				(n=1000)	(n=1)	(n=5)
rb1	–	–	24.04	24.04	23.2	**22**
rb2	–	10	8.2	7.25	**6.5**	**6.5**
rb3	–	–	22.4	17.9	17.5	**16.7**
rb4	–	–	**16.9**	**16.9**	**16.9**	**16.9**
rb5	–	–	24.54	24.54	**21.44**	**21.44**
hok	–	–	28.5	29.66	20.7	**20.1**
SL	11.80	–	13	**12.9**	13.0	**12.9**
attenuator	8.3	–	8.7	8.6	8.7	**8.5**
s15	6.60	–	7.1	**6.6**	7.1	7.1
sbox leader	–	7.9	**5.2**	**5.2**	**5.2**	**5.2**
thiM leader	–	–	16.13	14.84	**12.3**	**12.3**
ms2	–	11.6	**6.6**	**6.6**	**6.6**	**6.6**
HDV	–	23.53	17.4	17.0	**16.8**	**16.8**
dsrA	8.0	–	8.3	8.2	**8.0**	**8.0**
ribD leader	–	–	10.71	**9.5**	**9.5**	**9.5**
amv	–	12.2	5.8	5.8	**5.74**	**5.74**
alpha operon	–	11.8	6.5	6.5	**6.1**	**6.1**
HIV-1 leader	–	14.3	9.3	11.3	**8.9**	**8.9**

3. For each base pair (i', j') in $\{stack_h - S_t\}$ that is compatible with S_t, place $\texttt{add}_{i',j'}$ to p after a_t using strategy $\mathbb{M}_4$.
4. For each base pair (i', j') in $\{stack_h - S_t\}$ that is incompatible with S_t,
   4.1. Find all the base pairs (i^*, j^*) in S_t that are incompatible with (i', j'), and ensure that each base pair (i^*, j^*) is deleted before the action $\texttt{add}_{i',j'}$.
   4.3. If a action $\texttt{del}_{i^*,j^*}$ appears in $\{a_{t+1}, \ldots, a_m\}$, move it up before $\texttt{add}_{i',j'}$ using strategy $\mathbb{M}_1$.
   4.4. Otherwise, introduce a pair of complementary actions $\texttt{del}_{i^*,j^*}$ and $\texttt{add}_{i^*,j^*}$ using strategy $\mathbb{M}_3$.

Using $\mathbb{M}_5$, we can introduce the simultaneous formation of a stack $stack_h$, which is incompatible with S_t, and destruction of existent stacks (or base pairs) that hamper the formation of $stack_h$. Since cooperative formation and destruction of stacks may contribute additional stacking energies for stabilizing the intermediate structures, better folding pathways with lower energy barriers may be rendered.

3. RESULTS

3.1 Benchmark tests

We benchmarked RNAEAPath against existing methods (BARRIEERS [9, 11], PathwayHunter [34], Findpath [10], and RNATabuPath [7]) by predicting low energy barrier folding pathways between two designated RNA secondary structures of 18 conformational switches. All the conformational switches were taken from the work of Dotu *et. al* [7]. Five of them are riboswitches, including rb1, rb2, rb3, rb4, and rb5. The metastable structures of these riboswitches have been experimentally determined by inline probing [19, 37]. The thirteen remaining cases concern conformational switches, including hok, SL (Spliced leader RNA), s15, s-box leader, thiM leader, ms2, HDV, dsrA, ribD leader, amv, alpha operon and HIV-1 leader. Sequences of these conformational switches can also be obtained from paRNAss web site [1], and some of the metastable secondary structures were computationally determined using RNAbor [12].

We summarize the results computed by PathwayHunter, the results computed by BARRIERS, the results computed by Findpath (with the look ahead parameter $k = 10$), the best results over 1000 runs found by RNATabuPath, and the best results over 1 run and 5 runs found by RNAEAPath in Table 1 respectively. And we use '$-$' to mark test cases that methods fail to apply to in the table. For all methods, free energies of the intermediate structures of the folding pathways (including PathwayHunter) are evaluated based on the Turner model using RNAeval (with -d1 option) from the Vienna RNA Package [15]. The default configuration parameters of RNAEAPath are as follows. MAX is 10, γ is 5, $\mathfrak{L}$ is 100, ℓ_1 is 10, ℓ_2 is 5 and ℓ_3 is 100. Due to the stochastic nature of the evolutionary algorithm, we report the best energy barrier of RNAEAPath found over both 1 run and 5 runs.

BARRIERS is the only exact solution that produces indirect pathways based on the Turner model. BARRIERS has already been compared with existing heuristic algorithms on the same test cases in the work of Dotu *et al.* [7]. We put the results of BARRIERS in the table just for the sake of comparison. It has been pointed out that BARRIERS gives provably globally optimal pathways in 4 out of 18 cases (i.e. SL, attenuator, s15 and dsrA). BARRIERS can not be directly applied to 5 cases because either the initial or the end structure is not locally optimal (i.e. rb2, sbox leader, ms2, amv and alpha operon), and can not converge in the remaining cases. Possibly due to the fact that both the number of RNA secondary conformations to consider and the computational resources required increase exponentially with the growing length of the RNA sequence and the growing range of energy barrier. PathwayHunter is an exact algorithm capable of producing the optimal direct folding pathways based on the Nussinov model. PathwayHunter can not be directly applied to 10 cases, because it requires the pair of input structures being able to form a 'pairwise-optimal' bipartite conflicting graph (see the work of Thachuk *et al.* [34] for details). It is not surprising that the performance of the exact algorithm, PathwayHunter, evaluated by free energy (in kcal/mol), is worse than the heuristic algorithms. This is because PathwayHunter is optimized based on the Nussinov model and only produces direct pathways, while the

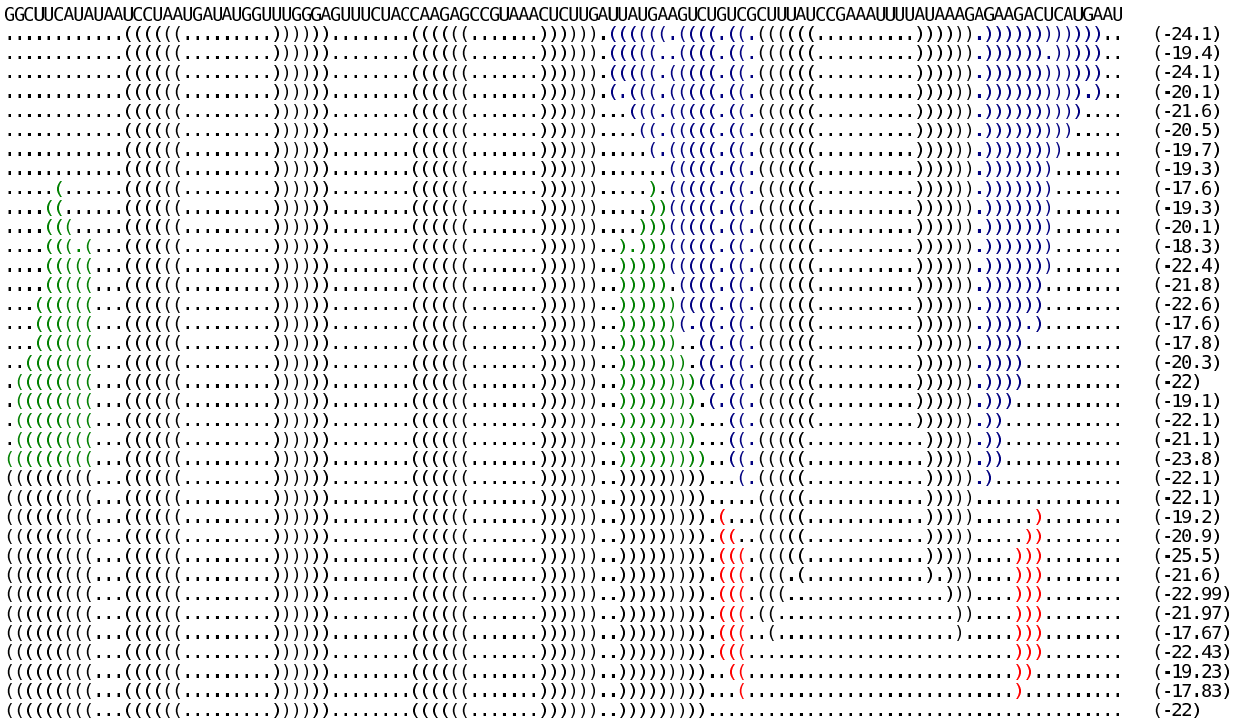

Figure 5: The near optimal indirect pathway between the two conformational secondary structures of the adenine riboswitch from *V. vulnificus* (rb2) predicted RNAEAPath.

optimal direct pathways predicted based on the Nussinov model may not be the optimal pathways (considering both direct and indirect pathways) based on the Turner model. All the remaining three methods are heuristics capable of producing both direct and indirect pathways based on the Turner model. Findpath produces folding pathways very quickly, however it performs worse than both RNATabuPath and RNAEAPath in most cases. RNATabuPath performs better than Findpath, but produces less optimal pathways than RNAEAPath. The energy barriers predicted by RNAEAPath over 5 runs are exactly the same as RNATabuPath in 5 cases, worse in 1 case, and better in all the remaining 12 cases.

Other heuristic algorithms (including a greedy algorithm of Voss *et al.* [36], a semi-greedy modification of the greedy algorithm, a greedy algorithm of Morgan, and Higgs [23] for predicting direct pathways and a variant of the Morgan-Higgs greedy algorithm capable of producing indirect pathways), that have been shown to perform considerably worse than RNATabuPath [7], are not listed.

By analyzing the best folding pathways produced by RNAEAPath, we found that most high-quality pathways involve the melting of stacks in the initial structure, the (possibly simultaneous) construction of stacks in the final structure, and the formation of auxiliary temporary stacks for obtaining folding pathways with lower energy barriers. We may take the lowest energy barrier folding pathway of rb2 found by RNAEAPath, shown in Figure 5, as an example. Some of the stacks in the initial structure (in blue) are gradually melting, while at the same time, an incompatible stack (in green) is being formed. The stack colored in red is an auxiliary temporary stack introducing intermediate structures with lower free energies. This example convinces us that the advantages of RNAEAPath mainly come from employing mutation strategies that guide the construction of folding pathways by the formation and destruction of stacks and introducing additional stacking interactions that are important for stabilizing the intermediate structures. Detailed low energy barrier folding pathways for all the test cases are available on RNAEAPath web site.

3.2 Control parameters and performance

In order to evaluate the performance of RNAEAPath with different parameter configurations, we played with several other control parameters, including ℓ_1, the number of top offsprings preserved in the next generation, varying from 1 to 16, ℓ_3, the size of population in each generation, varying from 80 to 120 and $\mathfrak{L}$, the total number of offsprings each individual is expected to produce, varying from 80 to 120. The detailed results are shown in supplementary data. In general, RNAEAPath produces pathways of roughly the same quality for most test cases with different control parameters, among which the default parameter setting is the best.

We explored the relationship between the performance of RNAEAPath and the number of generations completed by plotting energy barriers of the best folding pathways produced by RNAEAPath with the default parameters in each generation, as shown in Figure 6. In general, the energy barriers decrease dramatically in the first one or two generations, and then the decrements slow down and finally plateau within 10 generations. For instance, in the case of rb3, the predicted energy barriers of folding pathways in the initial population is 27.3 kcal/mol. It decreases by 7.2 kcal/mol (24.9%) through the first two generations and decreases by 2.5 kcal/mol (9.2%) through the next three generations. Through all the remaining generations, no further improvement is made.

We also evaluated the execution time for each run of RNAEAPath. All the tests were performed on a 32 bit PC with 2.4 GHz Quad-processor and 3.2 GB memory, running Fedora 11. With the

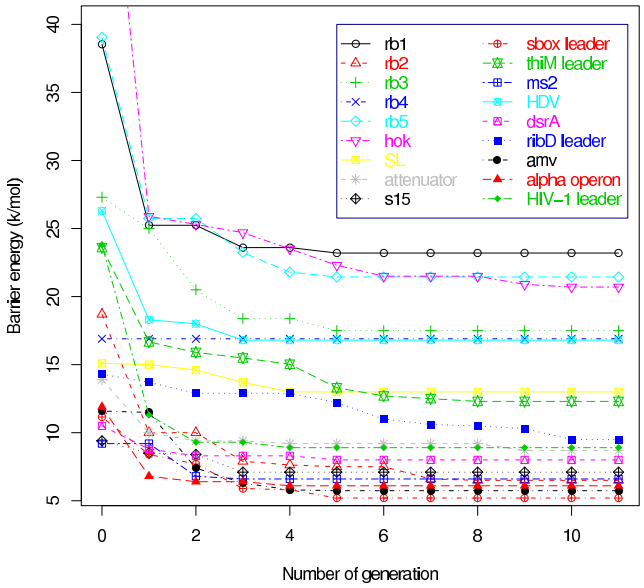

Figure 6: Energy barriers (in kcal/mol) of the best folding pathways of 18 conformational switches by the end of each generation during a typical run of RNAEAPath using the default parameters.

default control parameters, RNAEAPath terminates in 1 minute in the best case (rb4), 445 minutes in the worst case (hok), and 43 minutes on average. The detailed running times are shown in supplementary data. We did not perform direct comparisons between the running time of RNATabuPath and that of RNAEAPath, since RNATabuPath is only accessible via web server.

4. CONCLUSION

In conclusion, we have presented a new algorithm, RNAEAPath, for predicting low energy barrier folding pathways between conformational structures. RNAEAPath guides the construction of folding pathways through the destruction and formation of RNA stacks using various types of mutation strategies, and integrates them in a well-established computational framework of evolutionary algorithm. These mutation strategies can help reduce the search space and make it easier to jump out of local optima. By analyzing the results, we confirmed that most of the best folding pathways involve the formation of auxiliary stacks, or involve the cooperative formation and disruption of incompatible stacks. The benchmarking results show that RNAEAPath outperforms the existing heuristics on most test cases. We believe that this is because the construction of folding pathways in RNAEAPath captures important biological findings.

5. REFERENCES

[1] paRNAss, `http://bibiserv.techfak.uni-bielefeld.de/parnass/examples.html`.

[2] V. Bafna, H. Tang, and S. Zhang. Consensus folding of unaligned RNA sequences revisited. *J. Comput. Biol.*, 13:283–295, Mar 2006.

[3] C. K. Biebricher, S. Diekmann, and R. Luce. Structural analysis of self-replicating RNA synthesized by Qbeta replicase. *J. Mol. Biol.*, 154:629–648, Feb 1982.

[4] C. K. Biebricher and R. Luce. In vitro recombination and terminal elongation of RNA by Q beta replicase. *EMBO J.*, 11:5129–5135, Dec 1992.

[5] S. Bogomolov, M. Mann, B. VoÃ§, A. Podelski, and R. Backofen. Shape-based barrier estimation for RNAs, 2010.

[6] B. Bukau. Regulation of the Escherichia coli heat-shock response. *Mol. Microbiol.*, 9:671–680, Aug 1993.

[7] I. Dotu, W. A. Lorenz, P. Van Hentenryck, and P. Clote. Computing folding pathways between RNA secondary structures. *Nucleic Acids Res.*, 38:1711–1722, Mar 2010.

[8] A. Eiben. Evolutionary computing: the most powerful problem solver in the universe? *Dutch Mathematical Archive*, 5:126–131, 2002.

[9] C. Flamm, W. Fontana, I. L. Hofacker, and P. Schuster. RNA folding at elementary step resolution. *RNA*, 6:325–338, Mar 2000.

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

[11] C. Flamm, I. L. Hofacker, P. Stadler, and M. Wolfinger. Barrier trees of degenerate landscapes. *Z. Phys. Chem.*, 216, 2002.

[12] E. Freyhult, V. Moulton, and P. Clote. Boltzmann probability of RNA structural neighbors and riboswitch detection. *Bioinformatics*, 23:2054–2062, Aug 2007.

[13] M. Geis, C. Flamm, M. T. Wolfinger, A. Tanzer, I. L. Hofacker, M. Middendorf, C. Mandl, P. F. Stadler, and C. Thurner. Folding kinetics of large RNAs. *J. Mol. Biol.*, 379:160–173, May 2008.

[14] K. Gerdes, A. P. Gultyaev, T. Franch, K. Pedersen, and N. D. Mikkelsen. Antisense RNA-regulated programmed cell death. *Annu. Rev. Genet.*, 31:1–31, 1997.

[15] I. L. Hofacker. Vienna RNA secondary structure server. *Nucleic Acids Res.*, 31:3429–3431, Jul 2003.

[16] J. Liphardt, B. Onoa, S. B. Smith, I. Tinoco, and C. Bustamante. Reversible unfolding of single RNA molecules by mechanical force. *Science*, 292:733–737, Apr 2001.

[17] R. Lorenz, C. Flamm, and I. L. Hofacker. 2D projections of RNA folding landscapes, 2009.

[18] E. M. Mahen, P. Y. Watson, J. W. Cottrell, and M. J. Fedor. mRNA secondary structures fold sequentially but exchange rapidly in vivo. *PLoS Biol.*, 8:e1000307, Feb 2010.

[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*, 113:577–586, May 2003.

[20] J. Manuch, C. Thachuk, L. Stacho, and A. Condon. Np completeness of the direct energy barrier problem without pseudoknots. In R. Deaton and A. Suyama, editors, *15th International Conference DNA Computing and Molecular Programming*, pages 106–115, Berlin, Heidelberg, 2009. Springer-Verlag.

[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.*, 288:911–940, May 1999.

[22] R. K. Montange and R. T. Batey. Riboswitches: emerging themes in RNA structure and function. *Annu. Rev. Biophys.*, 37:117–133, 2008.

[23] S. Morgan and P. Higgs. Barrier heights between ground states in a model of rna secondary structure. *J. Phys. A*,

31(14):3153, 1998.

[24] F. Narberhaus. mRNA-mediated detection of environmental conditions. *Arch. Microbiol.*, 178:404–410, Dec 2002.

[25] F. Narberhaus, T. Waldminghaus, and S. Chowdhury. RNA thermometers. *FEMS Microbiol. Rev.*, 30:3–16, Jan 2006.

[26] W. Ndifon. A complex adaptive systems approach to the kinetic folding of RNA. *BioSystems*, 82:257–265, Dec 2005.

[27] R. Nussinov and A. B. Jacobson. Fast algorithm for predicting the secondary structure of single-stranded RNA. *Proc. Natl. Acad. Sci. U.S.A.*, 77:6309–6313, Nov 1980.

[28] R. Nussinov, G. Pieczenik, J. R. Griggs, and D. J. Kleitman. Algorithms for loop matchings. *SIAM Journal on Applied Mathematics*, 35:68–82, July 1978.

[29] A. Roth and R. R. Breaker. The structural and functional diversity of metabolite-binding riboswitches. *Annu. Rev. Biochem.*, 78:305–334, 2009.

[30] I. Shcherbakova, S. Mitra, A. Laederach, and M. Brenowitz. Energy barriers, pathways, and dynamics during folding of large, multidomain RNAs. *Curr Opin Chem Biol*, 12:655–666, Dec 2008.

[31] A. E. Simon and L. Gehrke. RNA conformational changes in the life cycles of RNA viruses, viroids, and virus-associated RNAs. *Biochim. Biophys. Acta*, 1789:571–583, 2009.

[32] D. J. Smith, C. C. Query, and M. M. Konarska. "Nought may endure but mutability": spliceosome dynamics and the regulation of splicing. *Mol. Cell*, 30:657–666, Jun 2008.

[33] J. Sponer, J. Leszczynski, and P. Hobza. Nature of nucleic acidâĹŠbase stacking: Nonempirical ab initio and empirical potential characterization of 10 stacked base dimers. comparison of stacked and h-bonded base pairs. *J. Phys. Chem.*, 100(13):5590–5596, 1996.

[34] C. Thachuk, J. Manuch, A. Rafiey, L. A. Mathieson, L. Stacho, and A. Condon. An algorithm for the energy barrier problem without pseudoknots and temporary arcs. *Pac Symp Biocomput.*, 15:108–119, 2010.

[35] D. H. Turner, N. Sugimoto, and S. M. Freier. RNA structure prediction. *Annu. Rev. Phys. Chem.*, 17:167–192, 1988.

[36] B. Voss, C. Meyer, and R. Giegerich. Evaluating the predictability of conformational switching in RNA. *Bioinformatics*, 20:1573–1582, Jul 2004.

[37] C. A. Wakeman, W. C. Winkler, and C. E. Dann. Structural features of metabolite-sensing riboswitches. *Trends Biochem. Sci.*, 32:415–424, Sep 2007.

[38] S. Wuchty, W. Fontana, I. L. Hofacker, and P. Schuster. Complete suboptimal folding of RNA and the stability of secondary structures. *Biopolymers*, 49:145–165, Feb 1999.

[39] P. Yakovchuk, E. Protozanova, and M. D. Frank-Kamenetskii. Base-stacking and base-pairing contributions into thermal stability of the DNA double helix. *Nucleic Acids Res.*, 34:564–574, 2006.

[40] P. Zhao, W. B. Zhang, and S. J. Chen. Predicting secondary structural folding kinetics for nucleic acids. *Biophys. J.*, 98:1617–1625, Apr 2010.