Design of Interacting Multi-stable Nucleic Acids for Molecular Information Processing

Effirul I. Ramlan and Klaus-Peter Zauner

Faculty of Science Computer and Information Technology,
University of Malaya, 50603 Kuala Lumpur, Malaysia
`effirul@um.edu.my`
School of Electronics and Computer Science
University of Southampton, SO17 1BJ, United Kingdom
`kpz@ecs.soton.ac.uk`

Abstract. Despite an exponential increase in computing power over the past decades, present information technology falls far short of expectations in areas such as cognitive systems and micro robotics. Organisms demonstrate that it is possible to implement information processing in a radically different way from what we have available in present technology, and that there are clear advantages from the perspective of power consumption, integration density, and real-time processing of ambiguous data. Accordingly, the question whether the current silicon substrate and associated computing paradigm is the most suitable approach to all types of computation has come to the fore. Macromolecular materials, so successfully employed by nature, possess uniquely promising properties as an alternate substrate for information processing. The two key features of macromolecules are their conformational dynamics and their self-assembly capabilities. The purposeful design of macromolecules capable of exploiting these features has proven to be a challenge, however, for some groups of molecules it is increasingly practicable. We here introduce an algorithm capable of designing groups self-assembling of nucleic acid molecules with multiple conformational states. Evaluation using natural and artificially designed nucleic acid molecules favours this algorithm significantly, as compared to the probabilistic approach. Furthermore, the thermodynamic properties of the generated candidates are within the same approximation as the customised trans-acting switching molecules reported in the laboratory.

1   Introduction

RNA molecules can possess multiple meta-stable conformations. An example from nature are riboswitches that directly bind to a specific metabolite, and harness the conformational changes in RNA to control gene expression (Mandal and Breaker, 2004; Nudler, 2006; Nudler and Mironov, 2004). From the perspective of constructing information processing nucleic acids (i.e., where a unit

is commonly known to switch from one conformation into another conformation), the design of sequences for RNA molecules with multi-stable conformations is important (Ramlan and Zauner, 2009). The development of a molecular switch that is triggered to alter its function in response to the binding of a small molecule demonstrates the possible application of this mechanism as a molecular information processing unit (Soukup and Breaker, 1999).

Secondary structure prediction of single RNA molecule is a classical problem in computational biology. This area of research has raised many interesting questions and propelled the emergence of various set of advanced prediction tools Reeder et al. (2006). One of the more common questions is the inverse prediction problem, where given an RNA structure, one needs to find a sequence that will fold into a desired target conformation. For RNA molecules, the number of such sequences always exceeds the number of structures. Thus, for a particular conformation of an RNA molecule, there are always many sequences that conform to it Schultes et al. (1998); Schuster (2006); Schuster et al. (1994). Random walks and stochastic approaches have been implemented to solve the sequence design problem as exemplified in *RNAinverse* from the *Vienna RNA package* Hofacker et al. (1994), *RNAdesigner* from the *RNAsoft* suite Andronescu et al. (2004) and *INFO-RNA* Busch and Backofen (2006). Commonly, these approaches rely on the singular folding prediction acting as a reference point during the optimisation process and explicitly avoid the occurrence of alternative conformations.

An alternate view towards the problem of secondary structure prediction was presented by Macke et al. (2001), whereby a search algorithm (called *RNAMotif*) was developed to find conversed RNA structural motifs that might exist given a set of RNA sequences. *RNAMotif* takes into consideration the assumption that certain structural motifs within a number of RNA structures are conserved within the same organisms or across a number of organisms. This approach also introduced a concept of "loose RNA structure hypothesis" in which a target structure description is represented as a tree with sub-trees corresponding to the different desired motifs. Using *RNAMotif*, Fogel et al. (2002); Lesnik et al. (2005) managed to show the relevance of conserved RNA structures over a wide range of organisms. This approach indirectly provide a different view of exploring the multiple meta-stable conformation characteristic of RNA molecule. For instance, "the loose RNA structure hypothesis" can provides us with the platform to describe a multi-stable RNA molecule with a number of conserved motifs that can later be translated into a set of RNA sequences given a pool of putative RNA candidates. In our context, this can be adopted into a filtering method towards validating RNA candidates with multiple meta-stable conformations.

Multi-stable conformation designers are required to generate RNA sequences that possess two or more meta-stable conformations with low free energies separated by a certain energy barrier that facilitates or hinders switching among them. The conformational switching, in this study, is triggered by the presence of a short oligonucleotide. RNA molecules with this switching mechanism are also known as trans-acting switches Nagel and Pleij (2002). In the laboratory, the conformational switching can also be achieved through the process of melt-

ing and rapid quenching of the molecule, illustrated by the multi-stable SV11 molecule Biebricher and Luce (1992); Zamora et al. (1995) in Fig. 1. Several
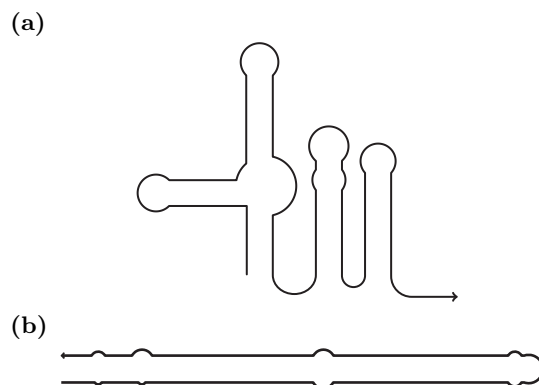
(a)

(b)

Fig. 1. The secondary structure representation of SV11 molecule as depicted in Ref. (Biebricher and Luce, 1992). The molecule has predominantly two confirmations. An active metastable RNA (a), an in vitro selected template for $Q\beta$ replicase, and a rod-like stable conformation (b) that no longer functions as the substrate in the replication process.

others naturally occurring RNA switches (*Tetrahymena* group I intron Pan and Woodson (1998); Russell and Herschlag (2001); Wu and Tinoco (1998), HDV ribozyme Chadalavada et al. (2000); Ferré-D'Amaré et al. (1998) and the *hok/sok* system of plasmid R1 from *E. coli* Franch et al. (1997); Gultyaev et al. (1997); Møller-Jensen et al. (2001)) follow the self-induced system Nagel and Pleij (2002) that does not require trans-acting factors (e.g., introduction of oligonucleotide) to initiate their conformational change.

At present, there is no designer for multiple molecules with multi-stable conformations that has been developed. There is, however, a multi-stable sequence designer for single molecule proposed by Flamm et al. (2001) and a software tool called *paRNAss* that has been developed to investigate the possibility of structural switching in RNA molecule by Geigerich et al. (1999). In this paper, we investigate the possibility of developing a sequence designer intended for multiple interacting molecules with multi-stable conformations using a deterministic approach. We then evaluate the performance of the deterministic algorithm against the proposed method of Flamm et al. (2001). To further evaluate the performance of the deterministic algorithm, we design nucleic acid sequences that can function as computational unit following a set of conformations from published materials and later simulate its thermodynamic properties against its reported sequences.

2 Identifying the Intersection Region for the Multi-stable Molecules

For a typical RNA molecule, approximately half of its conformation consists of multiple stretches of unpaired bases that can be identified as secondary structure loop motifs. In addition to hairpin loop, internal loop, bulges and multibranch loop, helices are also classified as secondary structure loop motifs. Helices are also known as stacked-loop. Thus, a stacked-loop is defined as a loop with size zero. Another element to consider is the pseudoknots. The presence of pseudoknots plays a major role in RNA self-splicing, translational autoregulation, and ribosomal frameshifting (t. Dam et al., 1992). However, little is known about the sequence dependence and thermodynamic stability of pseudoknots structures. Accordingly, it is excluded by the majority of the secondary structure prediction tools. There are a few algorithms that include pseudoknots in secondary structure prediction (Dirks and Pierce, 2003; Rivas and Eddy, 1999) using thermodynamic approximation and kinetic method, albeit at the cost of increased algorithm complexity. In shorter RNA sequences, it is rare to find the pseudoknot motif. For the purpose of designing computational nucleic acids the length of each molecule are restricted to 200 nt Ramlan and Zauner (2009). Considering the design space and the overhead (in terms of processing time) of predicting secondary structure with pseudoknots (Dirks and Pierce, 2003), we have decided to omit the pseudoknot motif in our design of the algorithm.

The design of an allosterically controlled ribozyme Penchovsky and Breaker (2005) with a single input receptor acting as regulator, requires two meta-stable conformations. For instance, a short nucleic acids strand is introduced to bind with a receptor site of the allosterically controlled ribozyme and triggers a conformational switch of the molecule Burke et al. (2002); Goodchild (1992); Komatsu et al. (2000); Penchovsky and Breaker (2005); Porta and Lizardi (1995); Wang et al. (2002). The two meta-stable conformations are the "inactive" conformation where a particular region of the molecule that is responsible for catalytic reaction (ribozyme core) is distorted and the "active" conformation which resembles the conformation of a catalytically active RNA. With the addition of multiple molecules in the designer, we also allow each molecule to have its own meta-stable conformations. This enables the sequence design algorithm to generate sequences for the design of a cascade of nucleic acid units where each unit has its own meta-stable conformations that affect other units (i.e., form external base pairs with other units) within the network.

For molecules with more than two meta-stable conformations, its sequence design follows the *Generalised Intersection Theorem* in Refs. (Schuster, 2006) and (Reidys et al., 1997). In order to cater for the multiple molecule aspect, we propose a new type of dependency graph Ψ^* illustrated in Fig. 2. The details of Fig. 2d are explained in Figs. 3 and 4.

In this dependency graph, the base position dependencies for both internal (meta-stable conformations for each molecule) and external (the inter-molecular pairing in each conformation for each molecule) base pairs are shown. In Fig. 2, we illustrate two random molecules (molecules I and II). Both molecules I and

II have three meta-stable conformations as illustrated in Fig. 2a–c. In state (b), the two molecules form external base pairs which is indicated by the dashed edges that connect two sets of vertices in the dependency graph. Instead of using the circular planar graph representation proposed by Flamm et al. (2001), we use the dot-plot representation Hofacker et al. (1994) to make paths and cycles more closely identifiable. This is particularly useful where the length of the molecules or the number of molecules and the number of meta-stable conformations overwhelm the circular planar graph representation proposed by Flamm et al. (2001).
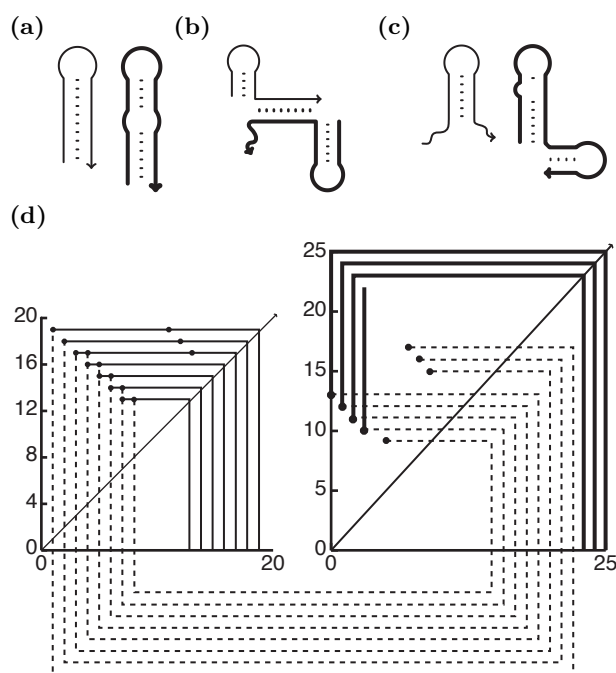

Fig. 2. The extended dependency graph Ψ^* depicting the base pairing interactions between two arbitrary multi-stable molecules. Both molecules I (thin) and II (bold) have three meta-stable conformations (a, b and c), for (a) and (c), these molecules fold internally to themselves and in (b), the two molecules externally bind together. In each state, the conformations are unique as visibly presented in the figure. The extended dependency graph (d) plots the base pairing relationship for both molecules separately into two dot-plot-like graphs. The base positions are represented in the x and y-axes, with dots denoting base pairs between node (x,y), plotted in the direction of 5′ to 3′, without repeating any previous base pairs. Paths are denoted by edges that connect these base pair points, while any edges that form a triangle path represent cycles (shown in Fig. 3b). The cross-linkage between the dot-plots (dashed) represents the inter-molecular binding between the two molecules. Refer to Figs. 3 and 4 for explanation of panel (d).

Similar to the dot-plot representation Hofacker et al. (1994), x-axis and y-axis represent the position in the sequence, in the usual 5' to the 3' order. We then plot all the occurring base pairs (in all states) in the structures only once, moving in the direction of the 5' to the 3' end. If two or more dots with the same values of x or y are present, a vertical line (for the y-axis) and a horizontal line (for the x-axis) are drawn to represent a path or a potential of a cycle formation (bold lines in Fig. 4a and b). These lines are then connected with the diagonal line. Any vertical line that is not connected to a horizontal line (in the upper half of the dot-plot, which is separated by the diagonal line) represents single (marked as "single" in Fig. 4a) which is a position that forms base pairs with the two different positions in each conformation. In order to find the start position of a path or cycle, we draw a vertical line from the intersection point with the diagonal line to the x-axis.
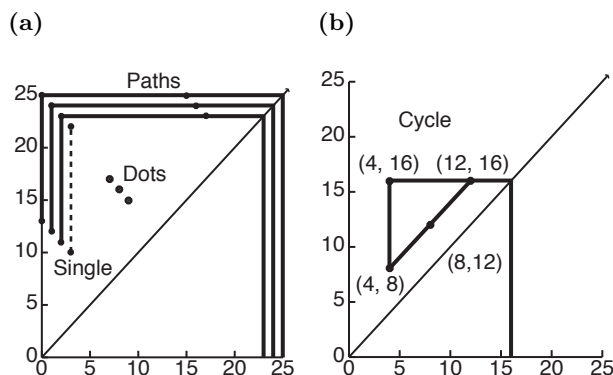

(a) **(b)**

Fig. 3. Elements of an extended dependency graph (Ψ^*). Path, singleton, and dot elements are presented in (a), while cycle element is shown in (b). A path is presented as black bullets connected by a vertical line and a horizontal line that connects to the diagonal line. A singleton is represented as two bullets that are connected with a vertical line. Dots are bullets which are not aligned with any other bullets (either vertically or horizontally). The coordinates in (b) are equivalent to the base pairs that form the cyclic dependency. In two conformations (a and b), position 4 can paired with position 8 and position 12 can paired with position 16 in conformation A, while position 4 and position 16, as well as position 8 and position 12 can form base pairs in conformation B.

For instance, in Fig. 4, we can trace a path that start from position 19 in molecule I and position 22 in molecule II as follows: 19 - 11 - 1 - 22$_{(II)}$. Cycles are identified, if in the upper half of the dot-plot, there exist a diagonal line that connects the start point of the vertical lines and the end point of the horizontal line forming a triangle (Fig. 4b). If there is an odd number of dots present in the triangular path, then one can conclude that the Ψ^* is not bipartite. If Ψ^*
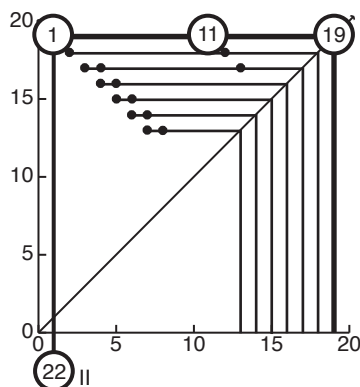
Fig. 4. Detail of the dependency graph shown in Fig. 2d. Path shown in bold, starts with position 19 which can pair with position 11 in one conformation, position 11 paired with position 1 in another conformation, and position 1 have the possibility of forming an external base pair with position 22 in molecule II, in another conformation.

is not bipartite, then there is no sequences that can be assigned to conform to the base pairing requirement of the cycle (cf. *Generalised Intersection Theorem* Reidys et al. (1997); Schuster (2006)). From the graph, we can also pick up dots (as "Dots" in Fig. 4a) which is a base position that forms a base pair with the same base position in all conformations. From the dot-plot-like dependency graph (presented here) a list of dependency pathways consisting of paths, cycles, singles, dots and unpaired bases is produced.

With the list of dependency pathways established, we can construct the trees of all possible base pairing combinations based on the length of the paths and cycles in the dependency pathways. Let us consider a simple case where all of the base positions will be unpaired, then the combination of all bases would be equivalent to a set of bases $\mathcal{A} = \{A,U,C,G\}$. Accordingly, for dots (a base pair occurring in all conformations), the possible base combinations are equivalent to the set of complementary base pairings $\mathcal{B} = \{CG,GC,AU,UA,GU,UG\}$. To assign the complete base combination for the path, we can combine any base pairs from the set of $\mathcal{B}$, but for a cycle, we take the formation of base pair between the start and end positions into consideration.

In Fig. 5, the process of deriving the complete base pair configuration of a cycle and a path 4 nt long is illustrated as the recursive process of adding bases from the complementary set $\mathcal{B}$ to the rooted tree. The possible base combinations for cycles are always less than the possible base combinations for paths because cycles, require pairing base between the start and end positions.
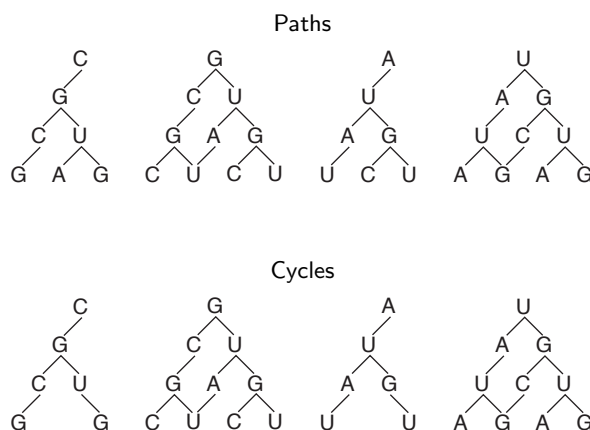
Paths

Cycles

Fig. 5. The combinatorial base configuration tree for a path and a cycle 4 nt long. For a base configuration starting with the bases `G` or `U`, the total number of configurations for both path and cycle are identical. However, for bases `C` or `A`, the end position can only be occupied by a complementary base to ensure that a cycle is formed.

3 Deterministic Approach for Solving the Interacting Multi-stable RNA molecules

A deterministic approach is proposed for optimisation based on two arguments: the design space that is restricted to a parameter envelope that includes all known computational nucleic acid units Ramlan and Zauner (2009) and, the multi-objective nature of the problem. For the desired computational nucleic acid units, the focus is not only on the total length of the molecules (200 nt) but equally important is the length restriction that is being apply to each of the secondary structure motif Ramlan and Zauner (2009). Because individual secondary structure elements are limited to the size of 25 nt, we can pre-compute the possible tree of base combinations to speed up the processing time.

The meta-stable conformations are known to exist within a certain energy gap from the minimum free energy of the molecule Schultes and Bartel (2000). By evaluating the free energy contribution of each possible base combination during base assignment, we can eliminate earlier a number of base combinations, and this makes the approach tractable for the design space of interest.

Given the list of dependency pathways, we consecutively select each element from the list and iteratively calculate the objective function ($\Xi(x)$) for each of the possible base assignments (x_i) to the element using a dynamic programming algorithm. For each element, we sort the possible base combination according to $\Xi(x_i)$ and retain the best candidates. The selection of element from the list of dependency pathways is made either randomly or in a descending order, i.e., long cycles are selected first, followed by long paths, singles, dots and unpaired

positions. After all elements in the dependency pathway list have been evaluated, a "traceback" routine Zuker et al. (1999) then retrieves the optimal solution. In

---

Algorithm 1 The algorithm for interacting multi-stable states molecules sequence design = *multiSrch*

1: generate set of pathways $\mathcal{D}(\Psi^*)$
2: initialise dummy sequence (x) with conserved bases and X's
3: add x into list of candidates L_c
4: **for** each pathway (p) in $\mathcal{D}(\Psi^*)$ **do**
5:     **for** each candidate sequence $x \in L_c$ **do**
6:         **if** p is cycle **then**
7:             Calculate $\Xi(x)$ for all cycle combinations where size $= l(p)$
8:             **if** $\Xi(x) < \Xi(X)$ **then**
9:                 replace $X = x$
10:             **else if** $\Xi(x) < \text{KEEP}$ **then**
11:                 add x to suboptimal list L_s
12:             **end if**
13:         **else if** p is path **then**
14:             Calculate $\Xi(x)$ for all cycle combinations where size $= l(p)$
15:             **if** $\Xi(x) < \Xi(X)$ **then**
16:                 replace $X = x$
17:             **else if** $\Xi(x) < \text{KEEP}$ **then**
18:                 add x to suboptimal list L_s
19:             **end if**
20:         **else if** p is unpaired **then**
21:             Calculate $\Xi(x)$ for assignment of $\forall p \in \{\texttt{A},\texttt{U},\texttt{C},\texttt{G}\}$
22:             **if** $\Xi(x) < \Xi(X)$ **then**
23:                 replace $X = x$
24:             **else if** $\Xi(x) < \text{KEEP}$ **then**
25:                 add x to suboptimal list L_s
26:             **end if**
27:         **end if**
28:     **end for**
29:     re-sort L_s according to $\Xi(x)$
30:     **for** each $x \in L_s$ **do**
31:         randomly select opponents
32:         compare x with random opponents
33:         re-sort L_s based on tournament score
34:     **end for**
35:     **for** each $x \in L_s$ until Λ **do**
36:         add x into L_c
37:     **end for**
38: **end for**

---

addition, we also retrieve suboptimal solutions where the value of the objective function is within a defined range from the minimal value. The size of the set of

suboptimal solution is denoted as (Λ). A large value of Λ would be desirable to improve the solution candidates, but significantly increases computation time.

In our algorithm, we also include a tournament ranking selection Bäck et al. (2000) to sort the suboptimal solutions during each iteration. This method decreases the bias of selecting only the best suboptimal solutions after each iteration. The algorithm, called *multiSrch*, is presented in detailed in Algorithm. 1. The KEEP variable denotes the error difference for the cost function between a partial candidate sequences against a minimum solution during each iteration, used to measure the selection of possible suboptimal candidate into a list (L_s).

The *multiSrch* algorithm calculates the objective function value for every possible base combination belonging to a given pathway element. At each stage, base combinations that contributed to the minimal objective function value are kept, together with base combinations with objective function values that are within the KEEP threshold. The process is repeated until all elements in the dependency pathways list have been selected. The algorithm prunes any possible base combinations that are not part of the suboptimal list. Using the values of the objective function calculated during the "fill" operation, we then apply a "traceback" routine to generate the set of sequences.
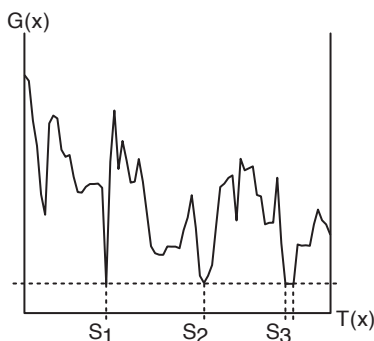

Fig. 6. A simplified representation of the free energy profile of a multi-stable RNA molecule with self-induced switching capability. Conceptually, for a molecule x, there is a time series of $\mathrm{T}(x)$ required for the molecule to kinetically fold into its native state. The $\mathrm{G}(x)$ axis represent the Gibbs free energy. The different peaks in the free energy axis indicate the formation and displacement of base pairs (structure re-arrangement) in order to arrive to its native conformation. The three lowest folds of S_1, S_2 and S_3 which are positioned almost at the same energy level represent the condition in which the three unique meta-stable states in a self-induced structural switching unit are normally present. The plots are simulated using *Kinfold* Flamm et al. (2000) for a random sequence with predicted folding equivalent to both conformations of the SV11 RNA species described in Ref. (Biebricher and Luce, 1992).

Two different objective functions are used by the algorithm, one for "self-induced switches" and the other for "trans-acting switches". The thermody-

namic characteristics of the two types of switches can be seen as sampling of the calculated free energy shown for the self-induced switch in Fig. 6 and for the trans-acting switch in Fig. 7. In figure 6, we observe that the free energy ΔG of the meta-stable conformations (S_1, S_2, and S_3) is relatively similar. This is consistent with the experimental finding of Schultes and Bartel (2000). Following Flamm et al. (2001), for self-induced switches, the objective function:

$$\Xi_A(x) = \Big(\sum_{i=1}^{n} E(x, S_i) \Big) - nG(x) \; + $$
$$\xi \sum_{i=1,j=1}^{n} \Big(E(x, S_i) - E(x, S_j) \Big)^2 \tag{1}$$

optimises the candidate sequences to favour the two meta-stable conformations by minimising the free energy differences between these meta-stable states to be close to or exactly zero (i.e., each state has the same free energy)
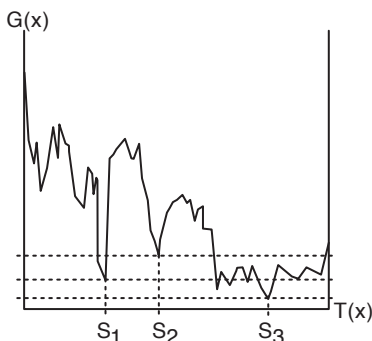

Fig. 7. A simplified representation of the free energy profile of a molecular PASS gate Penchovsky and Breaker (2005) with trans-acting structural switching capability. The schematic is produced with free energy calculation, simulated from *Kinfold* with estimated inter-binding energies by *RNAup* Mückstein et al. (2006), under the assumption that an RNA effector molecule is chosen to trigger the conformational change. Fold S_1 corresponds to an inactive gate, with fold S_2 representing the meta-stable state where refolding occur (as the effector molecule is introduced) and later form the active conformation in S_3. The three conformations differed in free energy.

Multi-stable trans-acting molecules are more suited for the design of nucleic acid computers. Control molecules (e.g., effector as input) can be introduced to activate or deactivate the function of these computers. From Fig. 7, we observe that the meta-stable states (S_1, S_2 and S_3) sit at a different energy level in contrast to the multi-stable state depicted in Fig. 6 above. The trans-acting switching molecule of an RNA PASS gate (described in Ref. (Penchovsky and Breaker, 2005)) simulated here requires the binding of an effector molecule to the

receptor site to trigger a conformational shift. In a time series of T(x) folding, calculated by *Kinfold* Flamm et al. (2000) the S_1 fold corresponds to the meta-stable state where the gate is inactive, fold S_2 denotes the meta-state where a short RNA effector molecule is introduced, and fold S_3 corresponds to the active conformation of the PASS gate. Following Flamm et al. (2001) the objective function:

$$\Xi_B(x) = \left(\sum_{i=1}^{n} E(x, S_i) \right) - nG(x) +$$
$$\xi \sum_{i=1,j=1}^{n} \left(E(x, S_i) - E(x, S_j) \right)^2 \qquad (2)$$

is suitable for the trans-acting switches, but computing the energy barriers between these meta-stable states to calculate the objective function is time consuming for the algorithm. A complete kinetic folding simulation is required for each possible base combination belonging to an element in the dependency pathway list. This significantly increases the CPU time of the algorithm.

To resolve this issue, we augment the objective function given in Eq. 1 with two additional terms. First is the energy gap value, (E_{gap}), that corresponds to the desired energy differences among meta-stable states, introduced by Penchovsky and Breaker (2005). The energy gap value is written as the difference of free energy between two meta-stable states given a sequence x, $E_{gap}(S_1, S_2) = |E(x, S_1) - E(x, S_2)|$. For more than two meta-stable states, the energy gap value is equivalent to:

$$E_{gap}(S_1, S_2, \ldots, S_m) = \sum_{i=1}^{m} \sum_{j=i+1}^{m} | E(x, S_i) - E(x, S_j) | \qquad (3)$$

where, m is the number of meta-stable states. The E_{gap} value can be tuned for specific applications. For the design of nucleic acid molecules for information processing, a value of $E_{gap} = 6$–10 kcal/mol is suggested in Ref. (Penchovsky and Breaker, 2005). The second term we introduce is the intermolecular base pairing efficiency, (ΔG_{int}) representing the disassociation of base pairs in the intramolecular binding region and formation of intermolecular base pairs calculated using the *RNAup* program Mückstein et al. (2006). With these extension to Eq. 1, the objective function is now:

$$\Xi_C(x) = \sum_{i=1}^{m} \left[E(x, S_i) + \Delta G_{int}(x, S_i) \right] - mG(x) +$$
$$\xi \sum_{i=1}^{m} \sum_{j=i+1}^{m} \left[E_{gap} - \left(E(x, S_i) - E(x, S_j) \right) \right]^2$$

$$(4)$$

where $E_{gap} = E_{gap}(x_1, x_2, \ldots, x_m)$ and the $\Delta G_{int}(x, S_i)$ is the free energy of meta-stable conformation S_i, where $\Delta G_{int}(x, S_i) = 0$, if S_i is does not have any intermolecular binding. We introduce another term (ΔG_{opt}) to allow one to specify target minimum free energy to accompany the target conformation. The objective function with ΔG_{opt} becomes:

$$\Xi_D(x) = \sum_{i=1}^{m} \Big[E(x, S_i) + \big(\Delta G_{opt} - \Delta G_{int}(x, S_i) \big) \Big]$$
$$- mG(x) + \xi \sum_{i=1}^{m} \sum_{j=i+1}^{m} \Big[E_{gap} - \big(E(x, S_i)$$
$$- E(x, S_j) \big) \Big]^2 \tag{5}$$

Equation 5 is the objective function we use for generating sequences where more than two molecular units are required.

From our trial runs, the implementation of weighting factor in the objective function failed to generate sequences that follow a specific criterion. In addition to the single aggregate objective function approach to handle the multi-objective optimisation required in the sequence design, a sorting-bins approach is derived for the algorithm. There are three different cost values that we need to consider. The first is the aggregate objective function itself (Eqs. 4 or 5), secondly the intermolecular base pairing efficiency measured as ΔG_{int} and thirdly the minimum free energy value. Each bin contains all candidates, but each bin is sorted according to the different cost function in descending order. In the algorithm, only the top candidates from each bin are selected for the next iteration. For specific design goals it can be advantageous to work with only one of the cost function. For example to design an effector molecule, optimising ΔG_{int} will increase the probability of generating sequences where the effector and receptor binding site (OBS) are perfectly complementary.

A tournament-ranking selection Bäck et al. (2000) is implemented instead of the straight forward sorting method to reduce the possibility of becoming stuck in local minima. This is necessary otherwise in the early stages of the execution, pre-mature pruning of base assignments would occurs. With only a fraction of the base positions already specified at this stage, the calculation of the objective function is not accurate. At every stage, in order to reduce the execution time of the algorithm, we need to prune any base combinations with poor objective function values. It is possible that these discarded base combinations could lead to the discovery of an optimal sequences. By starting with the longest path or cycle, we try to increase the number of bases available during the objective function calculation. The tournament-ranking selection adds to the effort of ensuring the list of suboptimal solution will be a sufficient representative of the complete base pair combinations in the early stages of the execution. In cases where different sequences for a given structure are required, the algorithm randomised the ordering of dependency pathway element. This can also be achieved by increasing the number of suboptimal candidate (Λ).

4   Evaluating the Performance of Multi-stable Sequence Designers

There are two types of structural switching (self-induced and trans-acting) that we need to consider in evaluating the multi-stable sequence design algorithms. The thermodynamic characteristics representative of the two switches are illustrated in Fig. 6 for self-induced switches and Fig. 7 for the trans-acting switches. In order to evaluate the overall performance of the multi-stable sequence designers, we examine the thermodynamic characteristic of both cases. Firstly the performance when designing self-induced switching molecules (where the free energy levels of the state are equal, i.e., $E_{gap} = 0$) and secondly, the performance of designing trans-acting molecules where the meta-stable states have different free energy levels but the difference E_{gap} is a design target.

In nature, only a few RNA molecules with multi-stable conformations have been identified. Most natural switches are either too large (more than 300 nt in length) or to complex (contain pseudoknot motifs) to be of interest for our purpose. Aside from the two SV11 molecules Biebricher and Luce (1992); Zamora et al. (1995) we have included a number of nucleic acid logic gates that have been engineered in the laboratory as part of the dataset for the comparison study. Although the selection of nucleic acid logic gates may seem inappropriate for self-induced structural switching because the logic gates follow the trans-acting switches, we use them because these molecules are known to have multi-stable conformations. Even the computational procedure that was used to design these nucleic acid logic gates initially ignored the influence of the effector molecules Penchovsky and Ackermann (2003); Penchovsky and Breaker (2005). Only later in the design process, the E_{gap} parameter is factored in to simulate the trans-acting structural switching.

The structures included in the dataset of RNA molecules with meta-stable conformations are listed in Tab. 1. The dataset is referred as "Multi-stable" (DS-MS). For the evaluation, we derived a multi-stable sequence designer named "*StochSrchMulti*", a variant of *RNAdesigner* Andronescu et al. (2004) which performed the best among existing single-state designers in the design space of interest Ramlan (2009). This variant algorithm is extended with the multi-stable conformation sequence design approached proposed by Flamm et al. (2001). The default parameters for *StochSrchMulti* are inherited from *RNAdesigner* Andronescu et al. (2004), while the local and non-local biases as proposed by Flamm et al. (2001) are tuned to the design space as described in Ramlan (2009).

In next two sections, we conduct two evaluation studies. First, *StochSrch-Multi* is compared against *multiSrch* to generate sequences for self-induces switches. This comparison study is referred as Test-Self-Switch (TSS). Secondly, *multiSrch* is evaluated to generate sequences for the trans-acting switches. This study is referred as Test-Trans-Switch (TTS).

Table 1. Multi-stable dataset (DS-MS) for the evaluation of designers for multi-stable molecules. Molecules numbered 1–11 have two meta-stable conformations, while the remaining molecules (12–18) have four meta-stable conformations. Only the two SV11 molecules are found in nature, while the remaining 16 have been constructed in the laboratory.

No.	Type	No. of states	No. of effectors	Ref.
1	SV11 plus	2	-	Biebricher and Luce (1992); Zamora et al. (1995)
2	SV11 minus	2	-	Biebricher and Luce (1992); Zamora et al. (1995)
3	RNA_PASS_1	2	1	Penchovsky and Breaker (2005)
4	RNA_PASS_2	2	1	Penchovsky and Breaker (2005)
5	RNA_NOT	2	1	Penchovsky and Breaker (2005)
6	RNA_OR	4	2	Penchovsky and Breaker (2005)
7	PORTA_PASS	2	1	Porta and Lizardi (1995)
8	BURKE_TRAP	2	1	Burke et al. (2002)
9	DNA_PASS_1	2	1	Stojanovic and Stefanovic (2003)
10	DNA_PASS_2	2	1	Stojanovic et al. (2002)
11	DNA_NOT	2	1	Stojanovic et al. (2002)
12	RNA_AND	4	2	Penchovsky and Breaker (2005)
13	DNA_AND_1	4	2	Stojanovic and Stefanovic (2003)
14	DNA_AND_2	4	2	Stojanovic et al. (2002)
15	DNA_AND_3	4	2	Kolpashchikov and Stojanovic (2005)
16	DNA_AND_4	4	2	Kolpashchikov and Stojanovic (2005)
17	DNA_MULTI_1	4	2	Stojanovic et al. (2002)
18	DNA_MULTI_2	4	2	Stojanovic and Stefanovic (2003)

4.1 Sequence Design for Self-induced Switches

In this section we evaluate the performance of *multiSrch* against *StochSrch-Multi* for the self-induced switches using dataset DS-MS (Tab. 1). The dataset comprises 18 molecules that have two or four meta-stable conformations. For the TSS study, we assigned loop elements to the region in the meta-stable conformations where external base pairing occurs. This is necessary because the majority of the molecules are engineered to switch in the presence of another molecule (trans-acting switches).

For each molecule in the datasets (DS-MS), a total of 100 runs are conducted for *StochSrchMulti*. In contrast to *StochSrchMulti*, *multiSrch* is a deterministic algorithm and therefore one run is sufficient. The TSS comparison study is conducted in two phase. In the first phase, only molecules from dataset DS-MS with two meta-stable conformations are compared (structures number 1 to 11). In the second phase the remaining structures 12 to 18, which are molecules with four multi-stable conformation are compared. The parameter settings for the *multiSrch* algorithm (cf. Algorithm 1) are shown in Tab. 2. The value of '$\mathcal{D}(\Psi^*)$ Order' denotes the type of sorting applied to the dependency pathways. Here the dependency pathways are sorted in descending order with regards to the number of edges on the path. For paths with an equal number of edges, the path starting earlier on the sequence takes precedence. The default values have been

Table 2. Default parameter setting for *multiSrch*. The Λ denotes the size of list for the suboptimal candidates. The $min(\Xi(x))$ term represents the minimum value of the objective function over all base assignments to one element of the dependency pathways.

Parameter	Value
$\mathcal{D}(\Psi^*)$ Order	descending
$\Xi(x)$	Eq.4 with $E_{gap} = 0$
Λ	50
KEEP	$\Xi(x) - min(\Xi(x)) = \pm 5.0$

optimised to suite the restricted search space of designing computational nucleic acids as investigated in Ramlan (2009).

For self-induced switches, we aim for an E_{gap} value of zero based on the free energy profile of self-induced switches illustrated in Fig. 6. The minimum free energy conformation is always a part of the meta-stable conformations. Therefore we can arrive at the sequence design of self-induced switches by minimising the free energy for any one of the meta-stable conformations and subsequently minimise the difference between the lowest free energy of a conformation against the other states.

To evaluate the quality of these sequences, we plotted the minimum free energy and the E_{gap} of each candidate into a bagplot graph Rousseeuw et al. (1999), shown in Fig. 8. Each panel shows the two dimensional boxplot of both the minimum free energy (MFE) of the secondary structures for the sequences and the energy gap between the two meta-stable conformations

The shaded grey and white regions of the graph are a representative of the sequences generated by *StochSrchMulti*. We then plot the data points of the sequences from the deterministic algorithm as ($\odot$). The lower left corner of each bagplots represents sequences with low free energy and low E_{gap}. Except for structures 5 and 7, sequences generated by *multiSrch* fall in the lower left corner. Generally the sequences of *multiSrch* (represented by $\odot$) are better than those generated by *StochSrchMulti*. An exception to this are the outliers in the plot of structure 10, where the outliers are of comparable in quality to the *multiSrch* sequences. This indicates that sequences from *multiSrch* more easily adhere to low energy gap. For structure 5, *StochSrchMulti* generates sequences with lower free energy from those generated by *multiSrch*, which have only a small E_{gap}. The opposite is observed in structure 7. Sequences from *multiSrch* have lower free energy but higher E_{gap} as compared to the sequences from *StochSrchMulti*. These two cases (structures 5 and 7) highlight the trade off between generating sequences with low E_{gap} and at the same time a low free energy, which are defined as two different terms in the objective function.

In Fig. 9, we can find data points representing sequences with low free energy and low E_{gap} (lower left corner of the bagplot) only for structure 12. The remaining bagplots either have a low E_{gap} with a high free energy or low free energy but with a high E_{gap}. Sequences generated by *multiSrch* for structure
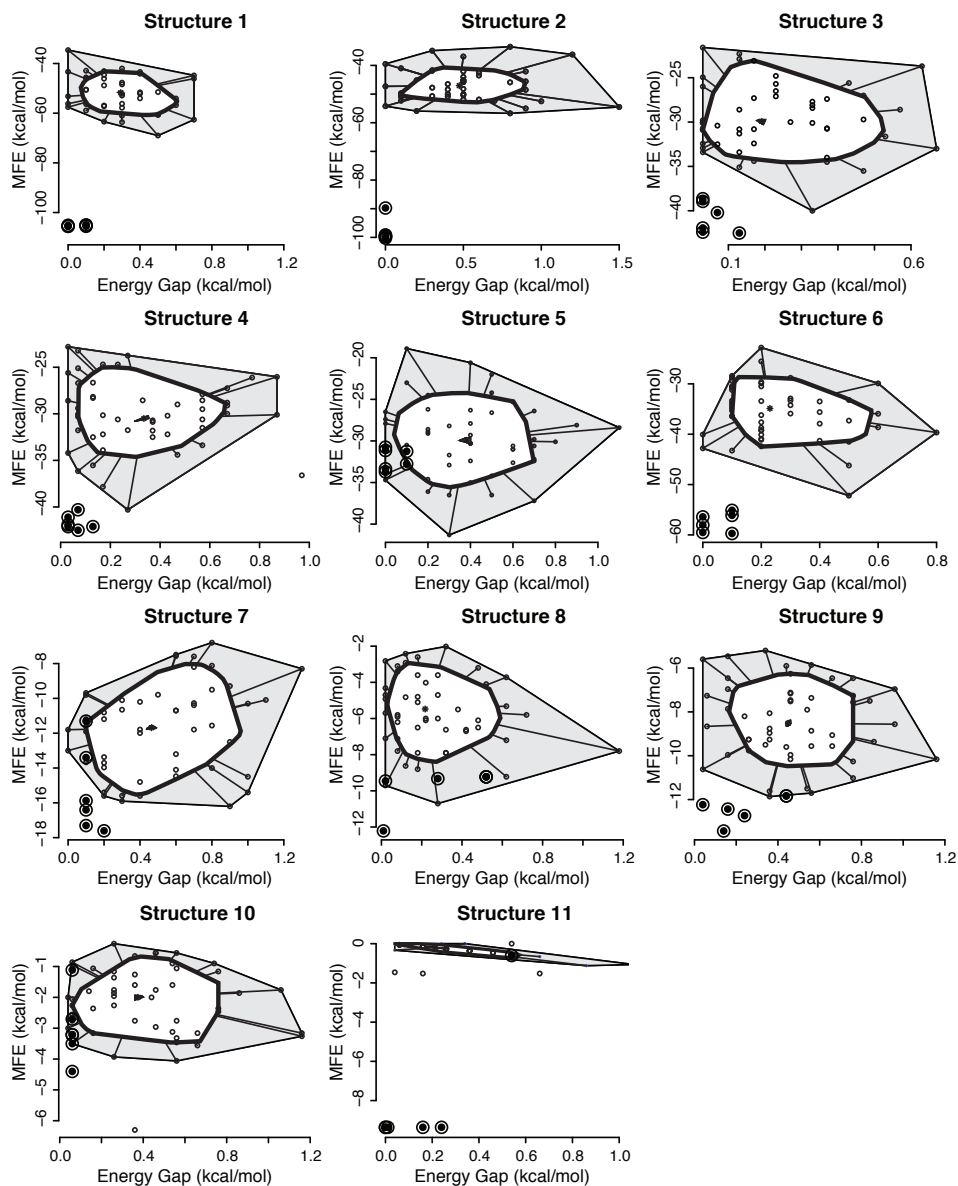
Fig. 8. The MFE and energy gap for the two-states molecules of DS-MS dataset (cf. Tab. 1). The main components of a bagplot graph are a *bag* that comprises of 50% of the data points (shaded grey), a *fence* that separates inliers from outliers and indicates the maximum and minimum data points equivalent to the whiskers in a univariate boxplot (white region). The depth median is represented by an asterisk (∗). The bagplot highlights the depth medium, the dispersion of data (size of *bag*), its correlation (indicated by the orientation of the *bag*, skewness (shape of the regions) and its tails (points at the boundary of the *loops* and outliers). Inliers and outliers are indicated by the black lines. Inliers are position inside the fence, while outliers are position outside the fence Text for details.
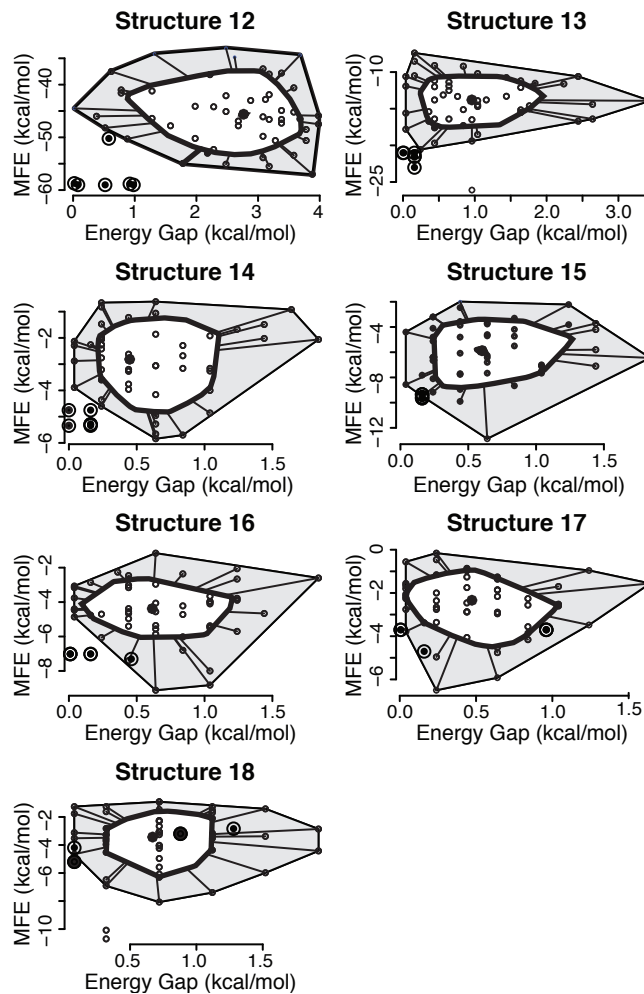
Fig. 9. The MFE and energy gap for the four states molecules in the DS-MS dataset (cf. Tab. 1). For each panel, the lower left corner represents an area where a low free energy and a low E_{gap} sequence is obtained. Only for structure 12 there are data points in the lower left corner belonging to sequences generated by *multiSrch*. The remainder of the plots either have a low E_{gap} with higher free energy or a low free energy with a higher E_{gap}. Refer to Fig. 8 and text for further details.

13, 14, 16 to 18 have a low E_{gap} and sequences generated by *StochSrchMulti* for structure 15 and 18 have a low E_{gap}. However the E_{gap} difference for sequences with low free energy in structure 13 to 18 is small ($E_{gap} < 1.0$ kcal/mol). If we compare only the free energy, then the sequences generated by *StochSrchMulti* have the lowest free energy for structure 13 to 15, 17 and 18.

As mentioned earlier, only a few natural occurring self-induced switching RNAs are known, and most are to long or to complex to be relevant for our design space. To evaluate the performance of *multiSrch* on a sufficiently large dataset DS-MS also includes molecular logic gates that have the advantage to have experimentally verified meta-stable conformations. However, these logic gates require the introduction of effector molecules in order to switch among their meta-stable conformations. To include these logic gate into the dataset for self-induced switches (TSS), we have substituted the external base pairing regions of the molecules with loops, thus leaving the majority of the meta-stable conformations to be loosely paired. For instance, structure number 12 to 16 that correspond to the nucleic acid AND gate would have a loosely paired meta-stable conformations because there exist two stretches of unpaired positions that initially (in trans-acting switches) should represent external base pairing formation with two effector molecules.

The lack of external base pairs has a direct affect for our deterministic algorithm. Pre-mature pruning of the possible base combination can occur, and this lowers the ability of the algorithm to produce sequences with lower free energy. The majority of the unpaired base positions are assigned with bases A or U, in an attempt to balance the free energy of each conformation. If one increases the probability of finding candidates with lower free energy, then the E_{gap} value would increase because of the loose pairing that usually would be bound with an effector. These conformations with loosely pairing are unlikely to form because the formation individual base pairs in RNA folding happens rapidly Tinoco and Bustamante (1999). Because our calculation of free energy for secondary structure folding is based on the nearest-neighbour model Xia et al. (1998), the free energy of base pair formation can be overwhelmed by the penalties the model applies to the unpaired bases.

The majority of these loosely paired conformations would yield significantly higher free energy (for instance, by definition, the free energy of an unpaired sequence is assumed to be 0.00 kcal/mol) and the E_{gap} between a meta-stable state where the majority of the base pairs are present against these loosely paired states can easily large. Similar to *multiSrch*, the *StochSrchMulti* tends to generate sequences with a slightly higher free energy to compromise for the increase of E_{gap} value. The comparison study for the self-induced switching molecule is intended to test the generality of our deterministic algorithm (*multiSrch*). The results show that the *multiSrch* algorithm managed to perform much better than *StochSrchMulti*. The results generated by *multiSrch* are well within the depth median of the candidates generated by *StochSrchMulti* (cf. bagplot Fig. 8 and 9). For the majority of the structures, *multiSrch* generates candidates with significantly smaller E_{gap} and lower free energy.

4.2   Sequence Design for Trans-acting Switches

The trans-acting switch test case (TTS), are important as they closely resemble the type of designs intended for nucleic acid information processing. Each of the molecules has multi-stable conformations that switch when a trigger molecule is

introduced. In order to evaluate the quality of sequences generated by *multiSrch* for trans-acting switches, we compiled a new dataset (DS-LG) comprised of the 14 nucleic acid logic gates from the previous DS-MS dataset. The complete meta-stable conformations (including external base pairing) are included in the DS-LG dataset instead of the loosely-paired conformation used in the DS-MS dataset.

The randomised *StochSrchMulti* algorithm is limited to design only a single multi-stable state molecule. Therefore, in order to conduct the evaluation, the results generated by *multiSrch* are directly compared with the actual experimental sequences from publications listed in Tab. 1. As discussed in Section 3, the *multiSrch* algorithm has been developed to generate sequences not only for multiple multi-stable molecules, but also the external base pairing formation that might occur between these molecules. Ideally, one would be able to construct a network of information processing units, consisting of computational nucleic acid units by running *multiSrch*. To support this, *multiSrch* must be able to handle the design of a single computational unit with two or three effector molecules effectively.

For trans-acting structural switching molecules, the E_{gap} value is optimised to be within a certain range, suited for the unfolding of existing base pairs and the formation of external base pair with the effector molecule(s). In order to test the two objective functions (Eqs. 4 and 5), we conduct two separate runs. For the first run, we kept the same parameter setting of *multiSrch* as described in Tab. 2 except for the size of the list of suboptimal candidates (Λ) from 50 to 100 and the target value for E_{gap} in the objective function (cf. Eq. 4) from 0.00 kcal/mol to 6.00 kcal/mol. The latter is the minimum of the range suggested by Penchovsky and Breaker (2005). We examined both the sequence and conformation characteristics of the 14 nucleic acid logic gates and compiled their thermodynamic properties in Tab. 3.

From the table, note the significant decrease in free energy between the published sequences and the sequences generated by *multiSrch*. Because the generated sequences are gradually built, during the early phase base assignment, the efficiency for external base pairing (ΔG_{int}) is too small to affect the calculation of the objective function (Eq. 4). Therefore the algorithm tends to optimise the free energy term, and thus discards potential candidates with high free energies. Only after a number of iterations, the regions that formed external base pairs will start to affect the value of the objective function. It is likely that at this stage, the base combinations that had previously been selected have already contributed to a low free energy, resulting in the significant decrease in free energy observed in Tab. 3. Accordingly, if the free energy of a meta-stable conformation is already low, then because of the dependency of the "intersection" regions, the E_{gap} for the subsequent meta-stable conformation might increase. Although anything between 10-15 kcal/mol remains within range recommended by Penchovsky and Breaker (2005). The energy gap is used for the objective function instead of the energy barrier among the meta-stable states because the simulation of a complete kinetic folding (Pair Kinetics and Helix Kinetic programs

Table 3. The thermodynamic characteristics of nucleic acids molecular gates, simulated by *RNAup* Mückstein et al. (2006) and the sequences generated by *multiSrch* with default parameter setting. The MFE value is representative of the inactive (OFF) conformation of the gates, while the Energy gap value is equal to the summation of the complete subtraction combination between the four states (cf. Eq.3). Mean of MFE and energy gap over 100 samples with errors for $\sigma = 1$. All the gates in the DS-LG dataset have been implemented, see Tab. 1 for references.

Type	Simulated		*multiSrch*	
	MFE	Energy Gap	MFE	Energy Gap
RNA_PASS_1	-35.80	8.60	-68.84±0.98	12.75±0.29
RNA_PASS_2	-26.97	8.94	-53.66±0.22	6.87±0.58
RNA_NOT	-37.77	6.74	-76.11±0.42	16.71±0.38
RNA_AND	-34.58	17.26	-66.19±0.18	16.32±0.70
RNA_OR	-53.56	17.45	-91.76±0.21	13.18±0.35
DNA_PASS_1	-12.82	0.63	-46.16±1.50	13.68±0.35
DNA_PASS_2	-12.80	0.84	-41.95±1.16	13.81±0.61
DNA_NOT	-16.33	2.41	-36.68±0.82	12.42±0.86
DNA_AND_1	-25.06	3.23	-91.96±0.22	17.22±0.96
DNA_AND_2	-19.01	0.54	-85.73±0.27	17.80±0.84
DNA_AND_3	-26.10	0.98	-100.40±0.29	19.89±0.66
DNA_AND_4	-28.07	4.90	-83.64±0.23	20.78±0.67
DNA_MULTI_1	-13.05	0.59	-80.41±0.34	25.85±0.69
DNA_MULTI_2	-42.36	3.28	-115.81±0.20	17.05±2.41

using Monte-Carlo simulation as suggested by Refs. (Higgs, 1993) and (Morgan and Higgs, 1998)) can increase the CPU time considerably.

For the second run, we substituted the objective function of *multiSrch* with Eq. 5 and changed the size of the pool of suboptimal candidates (Λ) from 100 to 50. In addition to the E_{gap}, it is necessary to supply the target value of ΔG_{opt}. In order to assign the suitable target values, starting from the rounded simulated values of the published sequence (cf. Tab. 3) for free energy and E_{gap}, we conduct trial runs with the size of suboptimal candidates (Λ) set to 5 and Eq. 5 as objective function, we gradually increase by 1 kcal/mol the value of ΔG_{opt} and E_{gap} to 5 kcal/mol, and also decrease these value by -1 kcal/mol until -5 kcal/mol. Based on the results of these trial runs, we select the target values for ΔG_{opt} and E_{gap} from the runs where the best set of sequences were generated. These values are listed in Tab. 4. The assignment of the target value for ΔG_{opt} and E_{gap} are intended to produce sequences that are closer in term of free energy and energy gap to the simulated values of the published sequences.

The result for the second run with objective function (Eq. 5) is listed in Tab. 4. When the target value for ΔG_{opt} and E_{gap} are specified *multiSrch* managed to generate results that are much closer in terms of MFE and E_{gap} to the published sequences. These improvements are significant when compared with the initial results listed in Tab. 3. In the default setting, the algorithm would search for the lowest free energy sequences that conform to the specified energy

Table 4. The thermodynamic characteristics for sequences generated by *multiSrch* using the objective function Eq.5 for the DS-LG dataset. The target values of both ΔG_{opt} and E_{gap} are approximation of the simulated Mückstein et al. (2006) values from published sequences (cf. Tab. 3). Mean of MFE and energy gap over 50 samples with errors for $\sigma = 1$

Type	Parameter value		*multiSrch*	
	ΔG_{opt}	E_{gap}	MFE	Energy Gap
RNA_PASS 1	-35.00	8.00	-36.46±0.53	10.38±1.08
RNA_PASS 2	-28.00	8.00	-29.23±0.78	10.78±0.73
RNA_NOT	-40.00	6.00	-42.22±0.95	8.16±0.23
RNA_AND	-35.00	17.00	-35.15±0.33	17.58±0.11
RNA_OR	-55.00	17.00	-57.06±1.49	18.57±0.34
DNA_PASS 1	-13.50	1.00	-14.07±0.07	1.00±0.07
DNA_PASS 2	-13.00	1.00	-14.41±0.83	1.19±0.62
DNA_NOT	-16.00	2.00	-17.19±0.52	2.03±0.05
DNA_AND 1	-30.00	3.50	-33.24±0.07	3.99±0.19
DNA_AND 2	-20.00	0.50	-21.39±0.08	0.83±0.09
DNA_AND 3	-27.00	1.00	-27.81±0.05	0.98±0.05
DNA_AND 4	-28.00	5.00	-30.97±0.31	5.43±0.10
DNA_MULTI 1	-15.00	0.50	-18.27±0.29	0.74±0.10
DNA_MULTI 2	-45.00	3.50	-45.97±0.44	3.90±0.10

gap value, and if such sequences exist, then it would also be possible to find sequences with higher free energy that still adhere to the energy gap target value. These sequences with higher free energy and marginally energy gap are readily available because *multiSrch* keeps any suboptimal candidates that are within ±5.00 kcal/mol of the minimal energy gap value. If one desires only sequences that comply to a certain minimum value of the objective function (e.g. KEEP $= \Xi(x) - min(\Xi(x)) = +2.0$), then we can directly restrict the suboptimal list to include only those candidates. The *multiSrch* algorithm only uses this target value as an additional criterion in pruning the base configuration branches. By specifying the exact values for the ΔG_{opt} and E_{gap} terms, the time required by the *multiSrch* algorithm would be further reduced because the amount of pruning for the base configuration branches has been increased. Any base combinations that already exceeds the amount of free energy and energy gap can be omitted. The specification of the exact values of the ΔG_{opt} and E_{gap} terms also allows one to increase the number of suboptimal candidates (Λ), because there are less branches to be evaluated as the number of pruning increase.

5 Conclusion

In the paper, we have introduced an algorithm for designing two types of molecular switching molecules related to the construction of nucleic acids information processor. From our review of the search space that is in accordance with the type of molecules suitable for the construction of information processor, we proposed

a deterministic approach to solve the problem of designing multiple multi-stable molecules. This deterministic algorithm is aimed not only for single multi-stable states molecule, but in addition the interaction between these multi-stable states molecule. Using dynamic programming, we are able to inspect and prune a complete base configurations assignment that are possible for each pathways. This pathway depict the complete dependency relationship of both internal and external base pairing. We developed a deterministic algorithm called *multiSrch* and evaluate the performance for both type of switching molecules against the random algorithm (*StochSrchMulti*). The findings for both types are in favour of our deterministic algorithm. Especially in the trans-acting case, where the ability to specifically designed sequences that confer to a particular MFE value and energy gap would be very useful in our task of designing nucleic acids information processors. Although, thus far the approach of combining the terms as a single objective seems reliable, perhaps a much better multi-objective optimisation approach can be applied to the deterministic algorithm to resolve the pre-mature pruning issue we observed in the self-induced structural switching case. The implementation of a more advance multi-objective optimisation strategy in our deterministic algorithm can greatly improve the processing time of the algorithm as the resorting element can be made more efficient. However, at the current state, our deterministic algorithm (*multiSrch*) have already shown a tremendous potential in generating nucleic acids candidates suitable for information processing.

Acknowledgments

Bibliography

Andronescu, M., Fejes, A. P., Hutter, F., Condon, A., Hoos, H. H., 2004. A new algorithm for RNA secondary structure design. Journal of Molecular Biology 336 (3), 607–624.

Bäck, T., Fogel, D. B., Michalewicz, T., 2000. Evolutionary Computation 1. Institute of Physics Publishing, Bristol and Philadelphia.

Biebricher, C. K., Luce, R., 1992. In vitro recombination and terminal elongation RNA by $Q\beta$ replicase. EMBO Journal 11 (13), 5129–5135.

Burke, D. H., Ozerova, N. D. S., Nilsen-Hamilton, M., 2002. Allosteric hammerhead ribozyme TRAPs. Biochemistry 41, 6588–6594.

Busch, A., Backofen, R., 2006. INFO-RNA—a fast approach to inverse RNA folding. Bioinformatics 22 (15), 1823–1831.

Chadalavada, D. M., Knudsen, S. M., Nakano, S., Bevilacqua, P. C., 2000. A role for upstream RNA structure in facilitating the catalytic fold of the genomic Hepatitis Delta Virus ribozyme. Journal of Molecular Biology 301, 349–367.

Dirks, R. M., Pierce, N. A., 2003. A partition function algorithm for nucleic acids secondary structure including pseudoknots. Journal of Computational Chemistry 24 (13), 1664–1677.

Ferré-D'Amaré, A. R., Zhou, K., Doudna, J. A., 1998. Crystal structure of a hepatitis delta virus ribozyme. Nature 395, 567–574.

Flamm, C., Fontana, W., Hofacker, I. L., Schuster, P., 2000. RNA folding at elementary step resolution. RNA 6, 325–338.

Flamm, C., Hofacker, I. L., Maurer-Stroh, S., Stadler, P. F., Zehl, M., 2001. Design of multistable RNA molecules. RNA 7, 254–265.

Fogel, G. B., Proto, V. W., Weekes, D. G., Fogel, D. B., Griffey, R. H., McNeil, J. A., Lesnik, E., Ecker, D. E., Sampath, R., 2002. Discovery of RNA structural elements using evolutionary computation. Nucleic Acids Research 30 (23), 5310–5317.

Franch, T., Gultyaev, A. P., Gerdes, K., 1997. Programmed cell death by *hok/sok* of plasmid R1: Processing at the *hok* mRNA 3-end triggers structural rearrangements that allow translation and antisense RNA binding. Journal of Molecular Biology 273 (1), 38–51.

Geigerich, R., Haase, D., Rehmsmeier, M., 1999. Prediction and visualization of structural switches in RNA. In: Altman, R. B., Dunker, A. K., Hunter, L., Klien, T. E. (Eds.), Proceedings of the Pacific Symposium on Biocomputing. Vol. 4. World Scientific Press, pp. 126–137.

Goodchild, J., 1992. Enhancement of ribozyme catalytic activity by a contiguous oligodeoxynucleotide (facilitator) and by 2'-O-methylation. Nucleic Acids Research 20 (17), 4607–4612.

Gultyaev, A. P., Franch, T., Gerdes, K., 1997. Programmed cell death by *hok/sok* of plasmid R1: Coupled nucleotide covariations reveal a phylogenetically conserved folding pathway in the *hok* family of mRNAs. Journal of Molecular Biology 273 (1), 26–37.

Higgs, P. G., 1993. RNA secondary structure: a comparison of real and random sequences. Journal de Physique I 3, 43–59.

Hofacker, I. L., Fontana, W., Stadler, P. F., Bonhoeffer, L. S., Tacker, M., Schuster, P., 1994. Fast folding and comparison of RNA secondary structures. Chemical Monthly 125 (2), 167–188.

Kolpashchikov, D. M., Stojanovic, M. S., 2005. Boolean control of aptamer binding states. Journal of the American Chemical Society 127, 11348–11351.

Komatsu, Y., Yamashita, S., Kazama, N., Nobuoka, K., Ohtsuka, E., 2000. Construction of new ribozymes requiring short regulator oligonucleotides as a cofactor. Journal of Molecular Biology 229, 1231–1243.

Lesnik, E. A., Fogel, G. B., Weekes, D., Henderson, T. J., Levene, H. B., Sampath, R., Ecker, D. J., 2005. Identification of conserved regulatory rna structures in prokaryotic metabolic pathway genes. Biosystems 80, 145–154.

Macke, T. J., Ecker, D. J., GUtell, R. R., Gautheret, D., Case, D. A., Sampath, R., 2001. RNAMotif, an RNA secondary structure definition and search algorithm. Nucleic Acids Research 29 (22), 4724–4735.

Mandal, M., Breaker, R. R., 2004. Gene regulation by riboswitches. Molecular cell biology 5, 451–463.

Møller-Jensen, J., Franch, T., Gerdes, K., 2001. Temporal translational control by a metastable RNA structure. Journal of Biological Chemistry 276 (38), 35707–35713.

Morgan, S. R., Higgs, P. G., 1998. Barrier heights between ground states in a model of RNA secondary structure. Journal of Physics A: Mathematical & General 31, 3153–3170.

Mückstein, U., Tafer, H., Hackermüller, J., Bernhard, S., Stadler, P. F., Hofacker, I. L., 2006. Thermodynamics of RNA-RNA binding. Bioinformatics 22 (10), 1177–1182, 10.1093/bioinformatics/btl024.
URL http://bioinformatics.oxfordjournals.org/cgi/content/ abstract/btl024v1

Nagel, J. H. A., Pleij, C. W. A., 2002. Self-induced structural switches in RNA. Biochimie 84, 913–923.

Nudler, E., 2006. Flipping riboswitches. Cell 126, 19–22.

Nudler, E., Mironov, A. S., 2004. The riboswitch control of bacterial metabolism. Trends in biochemical sciences 29 (1), 11–17.

Pan, J., Woodson, S. A., 1998. Folding intermediates of a self-splicing RNA: mispairing of the catalytic core. Journal of Molecular Biology 280 (4), 597–609.

Penchovsky, R., Ackermann, J., 2003. DNA library design for molecular computation. Journal of Computational Biology 10 (2), 215–229.

Penchovsky, R., Breaker, R. R., 2005. Computational design and experimental validation of oligonucleotide-sensing allosteric ribozymes. Nature Biotechnology 23 (11), 1424–1433.

Porta, H., Lizardi, P. M., 1995. An allosteric hammherhead ribozyme. Bio/Technology 13, 161–164.

Ramlan, E. I., 2009. Functional nucleic acids as substrate for information processing. Ph.D. thesis, University of Southampton, Southampton.

Ramlan, E. I., Zauner, K.-P., 2009. Nucleic acid enzymes: The fusion of self-assembly and conformational computing. International Journal of Unconventional Computing 5 (2), 165–189.

Reeder, J., Höchsmann, M., Rehmsmeier, M., Voss, B., Giegerich, R., 2006. Beyond Mfold: recent advances in RNA bioinformatics. Journal of Biotechnology 124, 41–55.

Reidys, C., Stadler, P. F., Schuster, P., 1997. Generic properties of combinatory maps: Neutral networks of RNA secondary structures. Bulletin of Mathematical Biology 59 (2), 339–397.

Rivas, E., Eddy, S. R., 1999. A dynamic programming algorithm for RNA structure prediction including pseudoknots. Journal of Molecular Biology 285 (5), 2053–2068.

Rousseeuw, P. J., Ruts, I., Tukey, J. W., 1999. The bagplot: a bivariate boxplot. The American Statistician 53, 382–387.

Russell, R., Herschlag, D., 2001. Probing the folding lanscape of the *Tetrahymena* ribozymel commitment to form the native conformation is late in the folding pathway. Journal of Molecular Biology 308, 839–851.

Schultes, E., Hraber, P. T., Labean, T. H., 1998. A parameterization of RNA sequence space. Complexity 4 (4), 61–71.

Schultes, E. A., Bartel, D. P., 2000. One sequence, two ribozymes: Implications for the emergence of new ribozyme folds. Science 289 (5478), 448–452.

Schuster, P., 2006. Prediction of RNA seconday structures: from theory to models and real molecules. Reports on Progress in Physics 69, 1419–1477.

Schuster, P., Fontana, W., Stadler, P. F., Hofacker, I. L., 1994. From sequence to shapes and back: A case study in RNA secondary structures. In: Proceedings of the Royal Society B: Biological Sciences. Vol. 255. pp. 279–284.

Soukup, G. A., Breaker, R. R., 1999. Nucleic acid molecular switches. Trends in Biotechnology 17, 469–476.

Stojanovic, M. N., Mitchell, T. E., Stefanovic, D., 2002. Deoxyribozyme-based logic gates. Journal of the American Chemical Society 124, 3555–3561.

Stojanovic, M. N., Stefanovic, D., 2003. A deoxyribozyme-based molecular automaton. Nature Biotechnology 21 (9), 1069–1074.

t. Dam, E., Pleij, K., Draper, D., 1992. Structural and functional aspects of RNA pseudoknots. Biochemistry 31 (47), 11665–11676.

Tinoco, Jr., I., Bustamante, C., 1999. How RNA folds. Journal of Molecular Biology 293, 271–281.

Wang, D. Y., Lai, H. Y., Feldman, A. R., Sen, D., 2002. A general approach for the use of oligonucleotide effectors to regulate the catalysis of RNA-cleaving ribozymes and DNAzymes. Nucleic Acids Research 30 (8), 1735–1742.

Wu, M., Tinoco, Jr., I., 1998. RNA folding causes secondary structure rearrangement. In: Proceedings of the National Academy of Sciences, USA. Vol. 95. pp. 11555–11560.

Xia, T., SantaLucia, Jr., J., Bunkard, M. E., Kierzek, R., Schroeder, S. J., Jiao, X., Cox, C., Turner, D. H., 1998. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. Biochemistry 37, 14719–14735.

Zamora, H., Luce, R., Biebricher, C. K., 1995. Design of artificial short-chained RNA species that are replicated by Qβ replicase. Biochemistry 34 (4), 1261–1266.

Zuker, M., Mathews, D. H., Turner, D. H., 1999. RNA Biochemistry and Biotechnology. NATO ASI. Kluwer Academic Publishers, Dordrecht, Ch. Algorithm and Thermodynamics for RNA Secondary Structure Prediction: A Practical Guide, pp. 11–43.