Predicting Model and Algorithm in RNA Folding Structure Including Pseudoknots

Zhendong Liu  
*School of Computer Science and Technology*  
*Shandong Jianzhu University*  
*Jinan 250101, P. R. China*  
*liuzd2000@126.com*

Daming Zhu  
*School of Computer Science and Technology*  
*Shandong University, Jinan 250101, P. R. China*  
*dmpzhu@sdu.edu.cn*

Qionghai Dai  
*Department of Automation, Tsinghua University*  
*Beijing 100084, P. R. China*  
*qudai@tsinghua.edu.cn*

Received 8 October 2017  
Accepted 22 February 2018  
Published 26 April 2018

The prediction of RNA structure with pseudoknots is a nondeterministic polynomial-time hard (NP-hard) problem; according to minimum free energy models and computational methods, we investigate the RNA-pseudoknotted structure. Our paper presents an efficient algorithm for predicting RNA structure with pseudoknots, and the algorithm takes $O(n^3)$ time and $O(n^2)$ space, the experimental tests in Rfam10.1 and PseudoBase indicate that the algorithm is more effective and precise. The predicting accuracy, the time complexity and space complexity outperform existing algorithms, such as Maximum Weight Matching (MWM) algorithm, PKNOTS algorithm and Inner Limiting Layer (ILM) algorithm, and the algorithm can predict arbitrary pseudoknots. And there exists a $1 + \varepsilon (\varepsilon > 0)$ polynomial time approximation scheme in searching maximum number of stackings, and we give the proof of the approximation scheme in RNA-pseudoknotted structure. We have improved several types of pseudoknots considered in RNA folding structure, and analyze their possible transitions between types of pseudoknots.

*Keywords*: RNA folding structure; algorithm; PTAS; pseudoknots; BHG.

This is an Open Access article published by World Scientific Publishing Company. It is distributed under the terms of the Creative Commons Attribution 4.0 (CC-BY) License. Further distribution of this work is permitted, provided the original work is properly cited.
1. Introduction

RNA is an important biomacromolecule which performs a wide range of functions in biological systems. RNA is a key component in several vital molecular biological processes. RNAs are three-dimensional molecules. The major driving force of RNA molecule is the set of base pairs of A–U, C–G match, and G–U mismatch; RNA can fold into a three-dimensional structure by forming base pairs; a pseudoknot is two overlapping base pairs; pseudoknots are known to exist in some RNAs; some RNA helices contain overlapping base pairs and more pseudoknots are used to refer pairs of substructures. RNA tertiary structure is a more stable structure, and RNA secondary structure prediction is the first step to predict RNA tertiary structure in RNA sequence.

Cross of base pairs form pseudoknots, and cross of stems can form pseudoknotted structure. Now it is difficult to compute large RNA molecules including pseudoknots for existing polynomial time-predicting algorithms. Finding the optimal RNA structures based on combination with stems has become the new method to predict RNA pseudoknotted structures. It is a nondeterministic polynomial-time hard (NP-hard) problem to find an optimal RNA secondary structure if we allow any set of base pairs, and some RNA structures are legal when base pairs obey the minimum separation requirement. For predicting secondary structures with pseudoknots, Nussinov has studied the case where the energy function is minimized when the number of base pairs is maximized, and has obtained an $O(n^3)$ time algorithm for predicting RNA secondary structures, but Nussinov algorithm cannot predict pseudoknotted structures. Algebraic dynamic programming algorithm for finding RNA-pseudoknotted structure with simple planar pseudoknots was proposed by Jens and Robert, the algorithm takes $O(n^4)$ time and $O(n^2)$ space. Efficient algorithm for finding optimal folding of an RNA structure has been firstly known by Michael Zuker; Pknots algorithm for RNA-pseudoknotted structure of predigesting model based on minimum free energy (MFE) has been presented by Rivas and Eddy, in which time complexity and space complexity are $O(n^6)$ and $O(n^4)$, respectively. The problem for predicting RNA secondary structure including pseudoknots is NP-complete, and maximizing the number of stacking pairs allowing pseudoknots in a planar secondary structure makes it NP-hard, so naturally people seek for approximation algorithms in the past. In mimic RNA structure, pseudoknots apparently exist. A heuristic algorithm including pseudoknots for finding RNA-pseudoknotted structures has been presented by Ren. Several publications show that extending the RNA structures including arbitrary pseudoknots indicates the problem of finding the optimum structure is NP-hard. People can find the more stable structure with arbitrary pseudoknots if RNA secondary structure is modeled by maximum weighted matching. The problem of the time and space complexities of predicting algorithms in the sparse case, sparse-related techniques have also been applied to RNA folding.

We analyze the RNA secondary structure. The contribution of this paper is to present an efficient algorithm for predicting RNA-pseudoknotted structure, where the time complexity of the algorithm is $O(n^3)$ and the space complexity is $O(n^2)$, and
we implement the algorithm in VC++ to complete the computation, “the experimental test in PseudoBase and Rfam10.1 of RNA database show that the algorithm is more effective and exact than other algorithms”, and the algorithm can predict arbitrary pseudoknots. Furthermore, we have proposed and proved that $1 + \varepsilon$ ($\varepsilon > 0$) polynomial time approximation scheme (PTAS) exists in searching maximum number of stackings. At last, we investigate the complexity of maximum number for stacking pairs of RNA structures and present a 2-approximation algorithm, analyzing the approximation ratio of the approximation algorithm. This paper also presents several types of pseudoknots considered in RNA folding structures, and analyzes their possible transitions between types of pseudoknots.

2. Predicting Model of RNA Structure
We discuss the model of RNA structure prediction with simple pseudoknots based on MFE through limiting pseudoknotted type.

2.1. Preliminary

(1) RNA secondary structure $S$: Let $S$ be a set of base pairs such as $s_i, s_j$, base $s_i$ or $s_j \in \{A, C, G, U\}$, $1 \leq i \leq n$.

(2) Pseudoknot: if $s_i, s_j$ and $s_i, s_{j'} \in S$, $i < i' < j < j'$, or $i' < i < j' < j$, then the RNA base sequence $s_i \ldots s_{i'} \ldots s_j s_{j'}$ composes a pseudoknot.

(3) Stem: the RNA structure closed by base pairs $(i, j)$ and $(k, l) \in S$, and $(i, j)$, $(I + 1, j - 1), \ldots, (k, l)$ are base pairs, $I \leq k < l \leq j$.

(4) $K$-stacking pairs: In the RNA secondary structures, we use $(s_i, s_{i+1}, \ldots, s_{i+k}; s_{j-k}, \ldots, s_{j-1}, s_j)$ to describe $k$ consecutive stacking pairs $(s_i, s_j)$, $(s_{i+1}, s_{j-1}); (s_{i+2}, s_{j-3}); \ldots; (s_{i+k-1}, s_{j-k}).$

(5) PTAS.

(6) BHG: basin hopping graph.

(7) Nested structure: if $s_i, s_j$ and $s_i, s_{j'} \in S$, $i < j < i' < j'$, or $i' < j' < i < j$, then the RNA base sequence $s_i, \ldots, s_{i'}, \ldots, s_j s_{j'}$ composes nested structure.

2.2. Algorithm of dynamic programming
Let $S = s_1 s_2, \ldots, s_n$, $s_{ij} = s_i, \ldots, s_j$, $V(i, j), W(i, j)$ and $W_M(i, j)$ can be computed as follows: $1 \leq i < j \leq n$.

In the formula, let $V(i, j)$ be the minimum energy for subsequence $s_{i,j}$ and $(s_i, s_j)$ form a base pair, let $W(i, j)$ be the minimum energy for subsequence $s_{i,j}$, the computation of $V(i, j)$ and $W(i, j)$ can be represented as Figs. 1 and 2. $E_1$ represents the energy of base pair, and $E_2$ represents the energy of stack. $EF$ represents the minimum energy of sequence $F$, $LF$ represents the length of sequence $F$, $W_M$ represents
the energy value of the multi-loop, and the computation of \( W(i, j) \) and \( W_M \) can be represented as Figs. 1 and 2. \( P \) represents the value of every base pair of the multi-loop, \( M \) represents the weight value of multi-loop, \( U \) represents the value of the length of base pairs, \( G_w \) represents the value of the pseudoknots. Based on principle of MFE, we use the thermodynamic parameters of M. Andronescu, A. Condon and Mathews,\(^2,3\) the algorithm can identify the pseudoknotted type of the query RNA structure, and output the RNA-pseudoknotted structure.

3. Predicting Algorithm of RNA Folding Structure

Based on principle of MFE, we design a polynomial time algorithm PreAlgorithm for predicting the RNA folding structure including pseudoknots. Stem is continuous stacking pairs in Fig. 3.
3.1. Definition

\[ E = mS_b + nP_b + lN_b + u + v, \]

where \( E \) is the total energy of an RNA-pseudoknotted structure, \( S_b \) is the total number of bases for RNA-pseudoknotted sequences, \( P_b \) is the total number of RNA pseudoknots, \( N_b \) is the number of the bases which do not match with others, and \( m \) is the weight of \( S_b \), \( n \) is the weight of \( P_b \), \( l \) is the weight of \( N_b \), \( u \) is the weight of coaxial stack, \( v \) is the weight of dangle bases. cf. Fig. 4.

\[ C(i, k: k + 1, j) \]

represents the bonus value of coaxial stack \( V(i, k) \) and \( V(k + 1, j) \), the dangle bases function \( L(i: I + 1, j) \) decided by the dangle bases \( s_i \) and base pair \((I + 1, j)\), the dangle bases function \( R(j: i, j - 1) \) decided by the dangle bases \( s_j \) and base pair \((i, j - 1)\).

\[ \min_{i < k < j - 1} \{ V(i, k) + V(k + 1, j) + C(i, k : k + 1, j) \} \]

\[ L(i: i + 1, j) + V(i + 1, j), R(j: i, j - 1) + V(i, j - 1) \]

\[ L(i: i + 1, j - 1) + R(j: i + 1, j - 1) + V(i + 1, j - 1) \]

(a)

(b)

Fig. 3. RNA folded structure.

Fig. 4. The representation (a) and formulas (b) of coaxial stack and dangle bases.
4. PTAS of RNA Folded Structure

RNA structure is used to any set of base pairs for an RNA molecule. cf. Fig. 3. Finding the optimal RNA folded structure based on combination with stems and loops has become new method to predict RNA-pseudoknotted structure, there are substructures in RNA secondary structure, for example, hairpin loop, internal loop, multibranched loop, bulge, stem and so on. In fact, RNAs are three-dimensional molecules. The set of base pairs in the three-dimensional structure of an RNA molecule is denoted by the secondary structure.

We design a PTAS according to the characteristic of stems in the RNA secondary structure, and analyze the optimized structure of stems. The paper divides the stem into several subsegments with the length $t$, and then searches the optimal structure formed by subsegments with the length less than $t$ as the approximation structure of given sequence.

Algorithm 1. PreAlgorithm $(S, E)$

1: Compute energy $E$ of RNA sequence $S$, $E = E_1 + E_2$
   $E_1$ is the energy value of RNA-nested structure, and $E_2$ is the energy value of RNA-pseudoknotted folding structure. Fill entry of $A[i, j]$ to save energy value of $E$ and stems.
2: $S_1 ← Ø, V_1 ← Ø, P_1 ← 0$
   Define $S_1$ to save the $k$-stems which are found in RNA structures, define $V_1$ to save the energy value in $m$-stems structure, and define $P_1$ to save the amount of RNA pseudoknots of $E_2$.
3: For $(k = 1; k \leq 42; k++)$ Search the prefix of $i$-stem $s_i$.
   For $(l = 1; l \leq 42; l++)$, search the suffix of $j$-stem $s_i$.
   Seek $k$-stem $s_k$ which is relative maximum of RNA molecule according to the MFE.
4: $v_k = \min \{v_1, v_2, \ldots, v_n\}$, $v_k$ is the energy value of $s_i$,
   $S_l ← S_l \cup \{s_i\}, V_l ← V_l \cup \{v_l\}$.
5: While ($f_d = 1$) do
   Mark the $i$-stem $s_i$ which is relative maximum in RNA structure $S$, $S ← S - \{s_i\}$.
6: Search the stem $s_j$ which is hypo-maximum in RNA structure $S$ except $s_i$
   which is relative maximum.
7: $S_l ← S_l \cup \{s_i\}$,
8: $V_1 ← V_1 \cup \{v_l\}$, until without $l$-stem in $S$.
9: $S_e = \{s_1, s_2, \ldots, s_n\}, V_0 = \{v_1, v_2, \ldots, v_n\}$.
10: Compute the number $P_1$ of pseudoknots in $S_l$.
11: Compute the energy values $E_1$ and $E_2$
12: Compute the energy $E$ according to $E_1$ and $E_2$.
13: Save and output $S, E$ and $P_1$.
4.1. Terminology

Let RNA sequence $S = s_1 s_2, \ldots, s_i, \ldots, s_n$, $s_i \in \{A, U, G, C\}$,

$$|S| = n, \quad 1 \leq i \leq n.$$ 

Given a stem $I$ in $S$, we suppose the RNA optimum structure $\text{OPT}(I) = \{x_1, x_2, \ldots, x_m\}$, $x_i$ is stem, and $x_i \in I, 1 \leq i \leq m$. cf. Fig. 5.

**Definition 1.** Let $L_{S[i,j]}$ be the length of stem $S[i,j]$ closed by base pairs $(i,j)$ and $(k,l) \in S, i \leq k < l \leq j$, then $L_{S[i,j]} = (k - i + 1)$ or $L_{S[i,j]} = (j - l + 1)$.

**Definition 2.** Let $\text{NPS}(S)$ be the number of stacking pairs in stem $S[i,j]$, then $\text{NPS}(S) = L_{S[i,j]}$.

**Lemma 1.** Let $S[i,j]$ be a stem of length $L_{S[i,j]}$ divide the stem $S[i,j]$ into $t$ segment: $s_1, s_2, \ldots, s_t$, let $L_{s_i}$ be the length of stem $1 \leq i \leq m$, then $L_{s_1} + L_{s_2} + \cdots + L_{s_t} = L_{S[i,j]}$, the number of stacking pairs after dividing the stem $S[i,j]$ is $\text{NPS}(S) = \sum L_{s_i} - m = L_{S[i,j]} - m$.

**Proof.** Let $(i,j)$ be external base pair in stem $S[i,j]$, without loss of generality, let $(k,l)$ be internal base pair in stem $S[i,j]$, according to the Definitions 1 and 2, then $L_{S[i,j]} = k - i + 1 = j - l + 1 = \text{NPS}(S[i,j]) + 1$.

**Similar.** $L_{s_i} = \text{NPS}(s_i) + 1, 1 \leq i \leq m$, then $\text{NPS}(s_i) = L_{s_i} - 1$, the number of stacking pairs after dividing the stem $S[i,j]$ into $m$ segment in stem $S[i,j]$ is $\text{NPS}(s_1) + \text{NPS}(s_2) + \cdots + \text{NPS}(s_t) = (L_{s_1} - 1) + (L_{s_2} - 1) + \cdots + (L_{s_t} - 1) = \sum L_{s_i} - t = L_{S[i,j]} - t$.

4.2. The analysis of PTAS

Given each $y_q \in \text{OPT}(I)$, let $L_q$ be the length of the stem $y_q$, $1 \leq L_q \leq n/2$, so the maximum number of stacking pairs for the stem $y_q$ is: $L_q - 1$.

Let $\text{MOP}(y_q)$ be the maximum number of stacking pairs for the stem $y_q$, then $\text{MOP}(y_q) = L_q - 1$. 

Fig. 5. The mRNA secondary structure.
If \( L_q > l(l > 2) \), we divide the stem \( y_q \) with the shorter stem of length \( l \) into \([L_q/l]\) segment.

We discuss with two cases as follows:

(a) If \([L_q/l] \neq L_q/l\), so \([L_q/l] = \text{int}(L_q/l) + 1\),

The reduced number of stacking pairs for primary stem \( x_q \) is \( \text{int}(L_q/l) + 1 \)

So \( \text{int}(L_q/l) + 1 < L_q/l + 1 \).

For any stem \( x_q \in \text{OPM}(I) \), the number of stacking pairs is more than \( L_q - 1 - (L_q/l + 1) \).

\[ L_q - 1 - (L_q/l + 1) = L_q - L_q/l. \]

Suppose the approximation scheme is \( \text{AS}(y_q) \)

So \( \text{AS}(y_q) > L_q - L_q/l. \)

Then

\[
\frac{\text{LOP}(y_q)}{\text{PT}(y_q)} < \frac{(L_q - 1)/(L_q - L_q/l)}{L_q/(L_q - L_q/l)} < 1/(1 - 1/l) = 1/(l - 1) = 1 + 1/(l - 1).
\]

That is, \( \text{MOP}(y_q)/\text{AS}(y_q) < 1 + 1/(l - 1) \)

So we get the results according to the first case:

\( \text{MOP}(y_q)/\text{AS}(y_q) < 1 + 1/(l - 1) \).

(b) If \([L_q/l] = L_q/l\), the reduced number of stacking pairs for stem \( y_q \) is \( L_q/l \)

The number of stacking pairs for the case is \( (L_q - 1)/L_q/l \)

Suppose the approximation scheme is \( \text{AS}(y_q) \)

So, \( \text{AS}(y_q) = (L_q - 1)/L_q/l \)

We compute the approximation ratio as follows:

\[
\frac{\text{MOP}(y_q)}{\text{AS}(y_q)} = \frac{(L_q - 1)/(L_q - 1) - (L_q/l)}{1/(1 - L_q/l)(L_q - 1))} < 1/(1 - 1/l - 1/l) = 1 + 2/(l - 2)
\]

So, \( \text{MOP}(y_q)/\text{AS}(y_q) < 1 + 2/(l - 2) \)
We get the results according to the second case:
\[ \text{MOP}(y_q)/\text{AS}(y_q) < 1 + 2/(l - 2), \quad (l > 2) \]

According to (a):
\[ \text{MOP}(y_q)/\text{AS}(y_q) < 1 + 1/(l - 1) \]

According to (b):
\[ \text{MOP}(y_q)/\text{AS}(y_q) < 1 + 2/(l - 2) \]
\[ 2/(l - 2) > 1/(l - 1) \]

According to two cases above, we have
\[ \text{MOP}(y_q)/\text{AS}(y_q) < 1 + 2/(l - 2) \]

Let \( \varepsilon = 2/(l - 2)(l > 2) \)

That is, \( \text{MGOP}(L_q)/\text{AS}(L_q) < 1 + \varepsilon \) (\( \varepsilon > 0 \))

So we can get the scheme to exist as \( 1 + \varepsilon(\varepsilon > 0) \) PTAS in searching maximum stacking pairs in RNA structure prediction.

5. Types of Pseudoknots Considered in RNA Folding Structure

If \( a, b \) are two local minima, then there exists a zigzag path connecting \( a \) and \( b \).

Reidys provides gfold software that can implement Boltzmann sampling. We can draw the RNA secondary structure with VARNA, and the RNA topological structures can be computed. We generalize the RNA-pseudoknotted framework based on BHG and study the differences in predicted folding behavior.

For a given RNA sequence \( S \), its energy landscape \( L \) is connected. For any base pairs given from secondary structures \( S_1 \) and \( S_2 \), there exists a path between \( S_1 \) and \( S_2 \), for any two local minimum \( m_1, m_2 \), then there exists a zigzag path connecting \( m_1 \) and \( m_3 \). We can define the path as follows: path \( P = (v_1, v_2, v_3, \ldots, v_k) \in L \), \( L \) is the energy landscape structure. If \( v_i < v_{i+1} = \cdots = v_{n-1} > v_n \), then for any structures \( v_{i+1} = \cdots = v_{n-1} \) are called peak points. If \( v_i > v_{i+1} = \cdots = v_{n-1} < v_n \), then for any structures \( v_{i+1} = \cdots = v_{n-1} \) are called valley points. If a path \( P \) fulfills three conditions: (1) max \( f(v_k) = S(x, y) \); (2) if \( v_i < v_{i+1} = \cdots = v_{n-1} > v_n \), then each \( v_m \) with \( i + 1 \leq m \leq n - 1 \) is a direct saddle separating the nearest valley points that the path \( P \) passed before and after \( v_m \); (3) if \( v_i > v_{i+1} = \cdots = v_{n-1} < v_n \), then each \( v_m \) with \( i + 1 \leq m \leq n - 1 \), there is a minimal shelf \( L \), we declare the path \( P \) is a zigzag path. \( P \) can be called BHG, then BHG is connected.

We can generalize the RNA structures with pseudoknots using the BHG and the sampling strategy for local minima. We can create a set for implementing the gradient walk of the class of pseudoknots, it comprises five types of pseudoknots as follows: Type S, Type H, Type K, Type L and Type M. cf. Fig. 6.
Type $S$ refers to structures without pseudoknots.

Removing base pairs is relatively simple since they will never result in an invalid structure, the general case involving five types of pseudoknots is rather involved, even with the restriction to structures, with at most one pseudoknot. See Table 1.

Adding base pairs is relatively simple since they will never result in an invalid structure, the general case involving five types of pseudoknots is rather involved, even with the restriction to structures, with at most one pseudoknot. See Table 2.

The paper presents an example named PKB92 of tobacco mild green mosaic virus, we investigate 27 bases with pseudoknots named PK1. Its structure can be correctly predicted with an energy of $-4.3$ kcal/mol by gfold. cf. Fig. 7.

Table 1. Possible transitions between types of pseudoknots upon removing a single base pair.

<table>
<thead>
<tr>
<th>Removing</th>
<th>M</th>
<th>L</th>
<th>K</th>
<th>H</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>L</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>K</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>H</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 2. Possible transitions between types of pseudoknots upon adding a single base pair.

<table>
<thead>
<tr>
<th>Adding</th>
<th>M</th>
<th>L</th>
<th>K</th>
<th>H</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>L</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>K</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>H</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>
The next pseudoknot-free MFE secondary is with an energy of 3.9 kcal/mol. cf. Fig. 8.

It is difficult to determine which base pairs can be added without changing the class of the RNA structure and to compute the changing result in energy without re-evaluating the structure. We restrict the subset of structures with H-type pseudoknots in restricted class cf. Fig. 9.

Fig. 7. Pseudoknots of PK1.

Fig. 8. Other pseudoknots of PK1.
According to the principle of the BHG and MFE, the paper provides a path-searching algorithm to connect the graph LM. We investigate the low-energy part of the BHG for PKB92 sequence, the PKB92 is more likely to fold the most stable secondary structure, and refold to form the pseudoknots. We label the LM by “Lx” and

![Diagram](image_url)

**Fig. 9.** (Continued)

![Diagram](image_url)

**Fig. 10.** The secondary structures with pseudoknots based on BHG.
label the saddles by “Sx”. The paper labels the edges by their energy barriers with [kcal/mol]. Optimal folding pathway of PKB92 sequence based on MFE, and the suboptimal pathways form the stem of pseudoknots which can be drawn in the BHG. Lx, Sx, Saddle points and local minima of the structures are as follows. cf. Fig. 10.

6. Experimental Comparison

The paper selected many sequences of Rfam10.1 and PseudoBase for experiment, and computes the energy value of stems including pseudoknots, according to the energy parameters.31,11,29 The polynomial time algorithm can compute RNA-nested structure and pseudoknotted structure in RNA sequences. We randomly picked the RNA subsequence in the Rfam10.1 and PseudoBase to compute the experiment.19 Many experiments on RNA-pseudoknotted structures indicated that the algorithm has above 83% predicting accuracy averagely. Our algorithm can predict more than 3300 bases of RNA sequences. The predicting accuracy of the algorithm descends with the increasing of the RNA base sequences. We have designed effective ways to improve the prediction accuracy for long sequences.

RNA sequence was generated with bases A,C,G,U, four experiments in family of PseudoBase can be computed in less than 45 seconds with quad-core CPU and 32 G}

<table>
<thead>
<tr>
<th>Sequence</th>
<th>PreAlgorithm Algorithm</th>
<th>Pknots Algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sensitivity</td>
<td>Specificity</td>
</tr>
<tr>
<td>Mottle virus</td>
<td>100%</td>
<td>98%</td>
</tr>
<tr>
<td>Mosaic virus</td>
<td>89%</td>
<td>97%</td>
</tr>
<tr>
<td>Ringspot virus</td>
<td>91%</td>
<td>95%</td>
</tr>
<tr>
<td>Sapiens mRNA</td>
<td>72%</td>
<td>81%</td>
</tr>
<tr>
<td>Time complexity</td>
<td>$O(n^3)$</td>
<td>$O(n^6)$</td>
</tr>
<tr>
<td>Space complexity</td>
<td>$O(n^2)$</td>
<td>$O(n^4)$</td>
</tr>
</tbody>
</table>

Table 4. Comparison of preAlgorithm algorithm and ILM, MWM and pknots.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>ILM</th>
<th>MWM</th>
<th>Pknots</th>
<th>PreAlgorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Sp%</td>
<td>Sn%</td>
<td>Sp%</td>
<td>Sn%</td>
</tr>
<tr>
<td>TYMV</td>
<td>82.9</td>
<td>100</td>
<td>63.3</td>
<td>100</td>
</tr>
<tr>
<td>Dr7741</td>
<td>96.8</td>
<td>94.5</td>
<td>84.5</td>
<td>91.2</td>
</tr>
<tr>
<td>HIV-1-RT</td>
<td>100</td>
<td>100</td>
<td>92.7</td>
<td>100</td>
</tr>
<tr>
<td>BMV1</td>
<td>84.6</td>
<td>85.4</td>
<td>85.4</td>
<td>87.8</td>
</tr>
<tr>
<td>TMV-3′-up</td>
<td>83.9</td>
<td>81.3</td>
<td>41.6</td>
<td>68.1</td>
</tr>
<tr>
<td>TMV-3′-down</td>
<td>68.5</td>
<td>76.6</td>
<td>51.1</td>
<td>73.6</td>
</tr>
<tr>
<td>HDV</td>
<td>82.5</td>
<td>100</td>
<td>45.3</td>
<td>67.9</td>
</tr>
<tr>
<td>Anti-HDV</td>
<td>66.8</td>
<td>100</td>
<td>38.5</td>
<td>70.9</td>
</tr>
<tr>
<td>Average</td>
<td>82.5</td>
<td>93.8</td>
<td>53.6</td>
<td>83.2</td>
</tr>
</tbody>
</table>

memory. We have performed experiments, and experiments show that accuracy of these experiments is valuable, the predicting accuracy outperforms existing algorithms, such as MWM algorithm, PKNOTS algorithm, ILM algorithm, etc. Evolutionary algorithm also provide a kind of important method in the RNA structure prediction. The structural alignment of RNA proves to be a useful computational technique for identifying noncoding RNA (ncRNA). The efficiency of our algorithm is faster than the other related algorithms in the RNA secondary structure and target structure (see Tables 3 and 4).

7. Conclusion and Future Work

In this paper, we have presented an efficient algorithm for predicting RNA structure with pseudoknots, and the algorithm takes $O(n^3)$ time and $O(n^2)$ space; the predicting accuracy, the time complexity and space complexity outperform existing algorithms, such as MWM algorithm, Pknots algorithm and ILM algorithm, and the algorithm can predict arbitrary pseudoknots. An $1 + \varepsilon$ ($\varepsilon > 0$) PTAS in searching maximum number of stackings has been presented, and we give the proof of the approximation scheme in RNA-pseudoknotted structure. We have improved several types of pseudoknots considered in RNA folding structure, and analyzed their possible transitions between types of pseudoknots.

It is a good computational method for characterizing the RNA folding structure using BHG. The RNA adopts an unexpected tandem three-way junction structure, and unspliced dimeric genomes are selected by the RNA conformer that directs packaging.

One of the strategies for feature selection that is often applied by brain–computer interface researchers is based on genetic algorithms. The fuzzy-rough feature selection and vaguely quantified rough set feature selection are coupled with CLONALG and AIRS for improved detection and computational efficiencies. Interactive segmentation of images has become an integral part of image-processing applications. Several graph-based segmentation techniques have been developed, which depend upon global minimization of the energy cost function. An adequate scheme of interactive segmentation still needs a skilled initialization of regions with user-defined seed pixels distributed over the entire image. Their algorithm idea is important for our paper in RNA structure prediction.

Acknowledgments

We are grateful to Lyngsø, Ieong Rivals, and Kucharík for their efficient work. Our work was supported by the National Natural Science Foundation of China under Grant Nos. 61672328 and 61272431. The research is also supported by Science and Technology Development Planning Project of Shandong Province (2014GGX101029), Science and Technology Project Plan of Ministry of Housing and Urban-Rural Development of P.R.C (2014-K8-071), Science and Technology Project.
Plan of Department of Housing and Urban-Rural Development of Shandong Province (KY003), Doctor Foundation of Shandong Jianzhu University (XNBS1438) and The Science and Research Plan of Luoyang branch of Henan tobacco Company (LYKJ201503).

References

Zhendong Liu is a visiting scholar in Department of Automation at Tsinghua University and he is a Professor, PhD, Master supervisor in School of Computer Science and Technology at Shandong Jianzhu University, China. He is a director, institute of computational biology. He received the MS degree in School of Computer Science and Technology from Shandong University, in 2005. He received the PhD degree in School of Computer Science and Technology from Shandong University, China, in 2014. His current research interests include algorithm and computational complexity, bioinformatics and computational biology, pattern recognition and artificial intelligence. He is a CCF senior member, ACM member and IEEE member.

Qionghai Dai is a Professor, Academician of China Engineering Academy, PhD supervisor, the Yangtze River scholar Professor in Department of Automation, Tsinghua University, Beijing, China. His research interests include computational photography and computational photography in bioinformatics.

Daming Zhu is a Professor, PhD supervisor in School of Computer Science and Technology at Shandong University, Jinan, China. He received the PhD degree in the Institute of Computing Technology from Chinese Academy of Sciences, Beijing, China, in 1999. His research interests include algorithm and computational complexity, bioinformatics and computational biology, pattern recognition and artificial intelligence.