MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **57** (2007) 157-168

ISSN 0340 - 6253

On 3D Graphical Representation of RNA Secondary Structure

Yusen Zhang[a]*
[a]*Department of mathematics, Shandong University at Weihai*
Weihai 264209, China

(Received May 15, 2006)

Abstract. A 3D graphical representation of RNA secondary structures using a three cartesian coordinates system has been derived for mathematical denotation of RNA structure. The 3D graphic representation resolves degeneracy completely and is mathematically proven to eliminate circuit formation. As an application, we make a comparison for 5 RNA pseudoknot sequences based on the new 3D graphic representation.

1 Introduction

RNA is a chain molecule, mathematically a string over a four letter alphabet. It is built from nucleotides containing the bases A(denine), C(ytosine), G(uanine), and U(racil). By folding back onto itself, an nucleic acids form structure, stabilized by hydrogen bonds between certain pairs of bases (A-U, C-G, G-U), and dense stacking of neighboring base pairs. The investigation of RNA secondary structures is a challenging task in molecular biology. RNA molecules have a large variety of functions in the cell which often depend on special structural properties. Current RNA secondary structure comparison algorithms have focused exclusively on tree structures owing to their relative simplicity for quantitative analysis[10-11]. But tree structures refer to mathematical constructs for RNA secondary structures without psedoknots. Here we should present a new representation to analyze and to compare RNA secondary structures with psedoknots. Recently, Liao et al. have proposed 3D, 6D and 7D graphical representations of RNA secondary structures[12-14], but the representation is not unique.

Here, we present a 3-dimensional graphical representation of RNA secondary structures, which has no circuit or degeneracy. The RNA psedoknots also can be represented as 3-dimensional graphical representations.

*Corresponding author:zhangys@sdu.edu.cn

2 3D graphical representation of RNA secondary structures

The secondary structure of an RNA is a set of free bases and base pairs forming hydrogen bonds between A-U, C-G and G-U. Let A', U', G', C' denote A, U, G, C in the base pairs A-U and C-G, respectively. Then we can obtain a special sequence representation of the secondary structure. We call it characteristic sequence of the secondary structure. For example, pseudoknot B corresponds the characteristic sequence

$$C'U'G'G'C'G'AUUGCG'A'G'A'C'C'A'UGUC'G'C'C'A'G'CUCU'G'G'U'C'U'C'CA$$

(from $3'$ to $5'$)(see Figure 1).

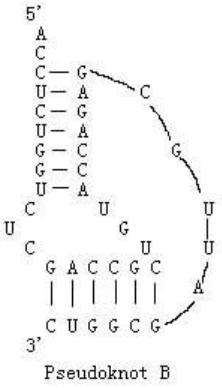

Figure 1: pseudoknot

We will illustrate the 3-dimensional characterization of RNA secondary structure. We construct a map between the bases of characteristic sequences and plots in 3D space, then we will obtain a 3D representation of the corresponding RNA secondary structure. In 3D space points, vectors and directions have three components, and we will assign the following basic elementary directions to the four free bases and two base pairs.

$(d_1\sqrt{u}, d_2\sqrt{u}, d_3\sqrt{u}) \longrightarrow A, (a_1v_1, a_2v_1, a_3v_1) \longrightarrow G,$

$(b_1v_2, b_2v_2, b_3v_2) \longrightarrow C, (c_1v_3, c_2v_3, c_3v_3) \longrightarrow U,$

$(d_1\sqrt{um}, d_2\sqrt{um}, d_3\sqrt{um}) \longrightarrow A', (a_1v_1\sqrt{m}, a_2v_1\sqrt{m}, a_3v_1\sqrt{m}) \longrightarrow G',$

$(b_1v_2\sqrt{m}, b_2v_2\sqrt{m}, b_3v_2\sqrt{m}) \longrightarrow C', (c_1v_3\sqrt{m}, c_2v_3\sqrt{m}, c_3v_3\sqrt{m}) \longrightarrow U',$

where

(1) u, w are different positive real numbers but not perfect square numbers.

(2) a_k, b_k and c_k $(k = 1, 2, 3)$ are real numbers.

(3) $s_k \neq 0, v_k \neq 0$ $(k = 1, 2, 3)$ are rational numbers.

$$(4)\ \begin{cases} d_1 = a_1 s_1 + b_1 s_2 + c_1 s_3, \\ d_2 = a_2 s_1 + b_2 s_2 + c_2 s_3, \\ d_3 = a_3 s_1 + b_3 s_2 + c_3 s_3. \end{cases} \quad \text{and} \quad \begin{bmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{bmatrix} \neq 0.$$

Thus we reduce a RNA secondary structure to a series of nodes $P_0, P_1, P_2, \ldots, P_N$, whose coordinates $x_i, y_i (i = 0, 1, 2, \ldots, N$, where N is the length of the RNA secondary structure being studied)satisfy

$$\begin{cases} x_i = d_1\sqrt{u}(A_i + \sqrt{w}A_i') + b_1 v_2(C_i + \sqrt{w}C_i') + c_1 v_3(U_i + \sqrt{w}U_i') + a_1 v_1(G_i + \sqrt{w}G_i') \\ y_i = d_2\sqrt{u}(A_i + \sqrt{w}A_i') + b_2 v_2(C_i + \sqrt{w}C_i') + c_2 v_3(U_i + \sqrt{w}U_i') + a_2 v_1(G_i + \sqrt{w}G_i') \\ z_i = d_3\sqrt{u}(A_i + \sqrt{w}A_i') + b_3 v_2(C_i + \sqrt{w}C_i') + c_3 v_3(U_i + \sqrt{w}U_i') + a_3 v_1(G_i + \sqrt{w}G_i') \end{cases} \quad (1)$$

where $A_i, A_i', C_i, C_i', G_i, G_i'$ and U_i, U_i' are the cumulative occurrence numbers of A, A', C, C', G, G', and U, U', respectively, in the subsequence from the 1st base to the i-th base in the sequence. We define $A_0 = C_0 = G_0 = U_0 = 0$, $A_0' = C_0' = G_0' = U_0' = 0$.

We call the corresponding plot set as characteristic plot set. The curve connecting all plots of the characteristic plot set is called 3DR-Curve (3D Curve of RNA), which is determined by condition (1) $\sim$ (4). In Figure 2, we show the curve that represent the pseudoknot B(see Figure 1).

Now we give two useful special cases.

1. Let $(\sqrt{u}, \sqrt{u}, \sqrt{u}) \longrightarrow A, (1,0,0) \longrightarrow G, (0,1,0) \longrightarrow C, (0,0,1) \longrightarrow T$,
$(\sqrt{uw}, \sqrt{uw}, \sqrt{uw}) \longrightarrow A', (\sqrt{w}, 0, 0) \longrightarrow G', (0, \sqrt{w}, 0) \longrightarrow C', (0, 0, \sqrt{w}) \longrightarrow T'$,
then we get the simple nondegeneracy 3D representation of RNA secondary structure:

$$\begin{cases} x_i = \sqrt{u}(A_i + \sqrt{w}A_i') + (G_i + \sqrt{w}G_i') \\ y_i = \sqrt{u}(A_i + \sqrt{w}A_i') + (C_i + \sqrt{w}C_i') \\ z_i = \sqrt{u}(A_i + \sqrt{w}A_i') + (U_i + \sqrt{w}U_i') \end{cases} \quad (2)$$

where u, w are positive real numbers, but not perfect square numbers.

2. Let $(d\sqrt{u}cos\theta, d\sqrt{u}sin\theta, d\sqrt{u}) \longrightarrow A, (v_1 cos\alpha, v_1 sin\alpha, v_1) \longrightarrow G$,
$(v_2 sin\alpha, v_2 sin\beta, v_2) \longrightarrow C, (v_3 cos\gamma, v_3 sin\gamma, v_3) \longrightarrow T$,
$(d\sqrt{uw}cos\theta, d\sqrt{uw}sin\theta, d\sqrt{uw}) \longrightarrow A', (v_1 cos\alpha\sqrt{w}, v_1 sin\alpha\sqrt{w}, v_1\sqrt{w}) \longrightarrow G'$,
$(v_2 sin\alpha\sqrt{w}, v_2 sin\beta\sqrt{w}, v_2\sqrt{w}) \longrightarrow C', (v_3 cos\gamma\sqrt{w}, v_3 sin\gamma\sqrt{w}, v_3\sqrt{w}) \longrightarrow T'$,
then we get a simple 3DR-Curve:

$$\begin{cases} x_i = d\sqrt{u}cos\theta(A_i + \sqrt{w}A_i') + v_2 cos\beta(C_i + \sqrt{w}C_i') + v_3 cos\gamma(U_i + \sqrt{w}U_i') + v_1 cos\alpha(G_i + \sqrt{w}G_i') \\ y_i = d\sqrt{u}sin\theta(A_i + \sqrt{w}A_i') + v_2 sin\beta(C_i + \sqrt{w}C_i') + v_3 sin\gamma(U_i + \sqrt{w}U_i') + v_1 sin\alpha(G_i + \sqrt{w}G_i') \\ z_i = d\sqrt{u}(A_i + \sqrt{w}A_i') + v_2(C_i + \sqrt{w}C_i') + v_3(U_i + \sqrt{w}U_i') + v_1(G_i + \sqrt{w}G_i') \end{cases} \quad (3)$$

where

(1) u, w are different positive real numbers, but not perfect square numbers.

(2) $s_k \neq 0, v_k \neq 0$ $(k = 1, 2, 3)$ are rational numbers.

(3)

$$s_1 = \frac{d(sin(\beta - \theta) + sin(\theta - \gamma) + sin(\gamma - \beta))}{sin(\beta - \alpha) + sin(\alpha - \gamma) + sin(\gamma - \beta)},$$

$$s_2 = \frac{d(sin(\theta-\alpha)+sin(\alpha-\gamma)+sin(\gamma-\theta))}{sin(\beta-\alpha)+sin(\alpha-\gamma)+sin(\gamma-\beta)},$$
$$s_3 = \frac{d(sin(\beta-\alpha)+sin(\alpha-\theta)+sin(\theta-\beta))}{sin(\beta-\alpha)+sin(\alpha-\gamma)+sin(\gamma-\beta)}.$$

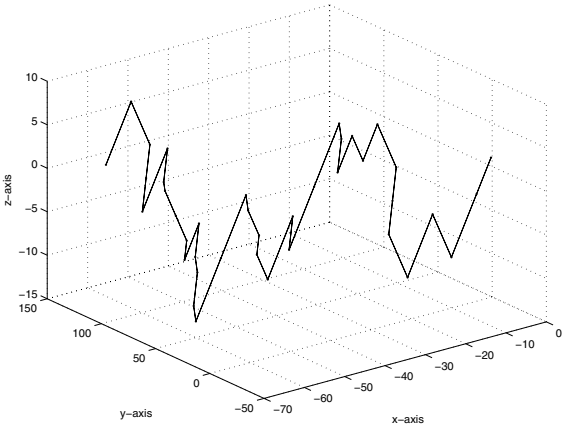

Figure 2: The 3DR-Curves of pseudoknot B

where Figure 2 corresponds to $v1 = 2; v2 = -4; v3 = 3; u = 1/3; w = 3; d = 2;$, $\alpha = \pi/3; \beta = -\pi/4; \gamma = 3\pi/4; \theta = -2\pi/3$ of case 2.

In next section, we will give some properties of our representation.

3 Properties

Property 1 For a given RNA secondary structure there is a unique 3DR-Curve corresponding to it.

Proof: Let x_i, y_i be the coordinates of the i-th base of RNA secondary structure, then we have

$$\begin{cases} x_i = d_1\sqrt{u}a_{1i} + b_1v_2c_{1i} + c_1v_3u_{1i} + a_1v_1g_{1i} \\ y_i = d_2\sqrt{u}a_{1i} + b_2v_2c_{1i} + c_2v_3u_{1i} + a_2v_1g_{1i} \\ z_i = d_3\sqrt{u}a_{1i} + b_3v_2c_{1i} + c_3v_3u_{1i} + a_3v_1g_{1i} \end{cases}$$

where $a_{1i} = A_{1i} + \sqrt{w}A'_{1i}$,$c_{1i} = C_{1i} + \sqrt{w}C'_{1i}$,$u_{1i} = U_{1i} + \sqrt{w}U'_{1i}$,$g_{1i} = G_{1i} + \sqrt{w}G'_{1i}$.

If (x_i, y_i, z_i) can also be expressed as

$$\begin{cases} x_i = d_1\sqrt{u}a_{2i} + b_1v_2c_{2i} + c_1v_3u_{2i} + a_1v_1g_{2i} \\ y_i = d_2\sqrt{u}a_{2i} + b_2v_2c_{2i} + c_2v_3u_{2i} + a_2v_1g_{2i} \\ z_i = d_3\sqrt{u}a_{2i} + b_3v_2c_{2i} + c_3v_3u_{2i} + a_3v_1g_{2i} \end{cases}$$

where $a_{2i} = A_{2i} + \sqrt{w}A'_{2i}$,$c_{2i} = C_{2i} + \sqrt{w}C'_{2i}$,$u_{2i} = U_{2i} + \sqrt{w}U'_{2i}$,$g_{2i} = G_{2i} + \sqrt{w}G'_{2i}$.

then, we have

$$\begin{cases} d_1\sqrt{u}(a_{1i}-a_{2i})+b_1v_2(c_{1i}-c_{2i}))+c_1v_3(u_{1i}-u_{2i})+a_1v_1(g_{1i}-g_{2i})=0 \\ d_2\sqrt{u}(a_{1i}-a_{2i})+b_2v_2(c_{1i}-c_{2i}))+c_2v_3(u_{1i}-u_{2i})+a_2v_1(g_{1i}-g_{2i})=0 \\ d_3\sqrt{u}(a_{1i}-a_{2i})+b_3v_2(c_{1i}-c_{2i}))+c_3v_3(u_{1i}-u_{2i})+a_3v_1(g_{1i}-g_{2i})=0 \end{cases}$$

By conditions (1), we can obtain

$$\begin{cases} s_1\sqrt{u}(a_{1i}-a_{2i})+v_1(g_{1i}-g_{2i})=0 \\ s_2\sqrt{u}(a_{1i}-a_{2i})+v_2(c_{1i}-c_{2i})=0 \\ s_3\sqrt{u}(a_{1i}-a_{2i})+v_3(u_{1i}-u_{2i})=0 \end{cases}$$

then, we get $a_{1i}=a_{2i}, c_{1i}=c_{2i}, u_{1i}=u_{2i}, g_{1i}=g_{2i}$.

So, $A_{1i}=A_{2i}, A'_{1i}=A'_{2i}, C_{1i}=C_{2i}, C'_{1i}=C'_{2i}, U_{1i}=U_{2i}, U'_{1i}=U'_{2i}, G_{1i}=G_{2i}, G'_{1i}=G'_{2i}$.

That means, for given x-projection, y-projection and z-projection of any point $P=(x,y,z)$ on 3DD-Curve, after uniquely determining the number a_p, g_p, c_p, u_p , a'_p, g'_p, c'_p and u'_p of A, G, C, U, A', G', C' and U' from the beginning of the sequence to the point P, by successive x-projection,y-projection and z-projection of points on the sequence, we can recover the original DNA sequence uniquely from the DNA graph.

The vector pointing to the point P_i from the origin O is denoted by r_i. The component of r_i, i. e. x_i and y_i are calculated by Eq. (1). Let $\Delta r_i = r_i - r_{i-1}$, then we have Property 2.

Property 2 For any $i=1,2,\ldots,N$, where N is the length of the studied RNA secondary structure, the vector Δr_i has at most eight possible direction.

Proof:Actually, the components of Δr_i,i.e., Δx_i and Δy_i can be calculated for each possible residue (A, G, C, U, A', G', C' and U') at the i-th position of the RNA secondary structure by using Eq.(1). For example, when the i-th residue is A, we find $\Delta x_i = d_1\sqrt{u}$, $\Delta y_i = d_2\sqrt{u}$ and $\Delta z_i = d_3\sqrt{u}$. This result is independent of the conformation state of the (i-1)-th residue. The eight vectors, for examples, $(d_1\sqrt{u}, d_2\sqrt{u}, d_3\sqrt{u})$ are called the direction of Δr_i. The direction numbers and the length of Δr_i for each possible residue type at the i-th position are summarized as follows(Table 1).

Table 1: eight possible directions

	Δx_i	Δy_i	Δz_i	$\lvert\Delta r_i\rvert$
A	$d_1\sqrt{u}$	$d_2\sqrt{u}$	$d_3\sqrt{u}$	$\sqrt{u}\sqrt{d_1^2+d_2^2+d_3^2}$
C	b_1v_2	b_2v_2	b_3v_2	$\lvert v_2\rvert\sqrt{b_1^2+b_2^2+b_3^2}$
G	a_1v_1	a_2v_1	a_3v_1	$\lvert v_1\rvert\sqrt{a_1^2+a_2^2+a_3^2}$
U	c_1v_3	c_2v_3	c_3v_3	$\lvert v_3\rvert\sqrt{c_1^2+c_2^2+c_3^2}$
A'	$d_1\sqrt{uw}$	$d_2\sqrt{uw}$	$d_3\sqrt{uw}$	$\sqrt{uw}\sqrt{d_1^2+d_2^2+d_3^2}$
C'	$b_1v_2\sqrt{w}$	$b_2v_2\sqrt{w}$	$b_3v_2\sqrt{w}$	$\lvert v_2\sqrt{w}\rvert\sqrt{b_1^2+b_2^2+b_3^2}$
G'	$a_1v_1\sqrt{w}$	$a_2v_1\sqrt{w}$	$a_3v_1\sqrt{w}$	$\lvert v_1\sqrt{w}\rvert\sqrt{a_1^2+a_2^2+a_3^2}$
U'	$c_1v_3\sqrt{w}$	c_2v_3	c_3v_3	$\lvert v_3\sqrt{w}\rvert\sqrt{c_1^2+c_2^2+c_3^2}$

Property 3 There is no circuit or degeneracy in our 3DR-Curve.

Proof: We assume that:(1) the number of nucleotide forming a circuit is e; (2) the number of A, G, C, U, A', G', C' and U' in a circuit is $a_e, g_e, c_e, u_e, a'_e, g'_e, c'_e$ and u'_e, respectively. So $a_e + g_e + c_e + u_e + a'_e + g'_e + c'_e + u'_e = e$. Because $a_e A, g_e G, c_e C, u_e U, a'_e A', g'_e G', c'_e C'$ and $u'_e U$ form a circuit, the following equation holds:

$$\begin{cases} d_1\sqrt{u}(a_e + \sqrt{w}a'_e) + b_1v_2(c_e + \sqrt{w}c'_e)) + c_1v_3(u_e + \sqrt{w}u'_e) + a_1v_1(g_e + \sqrt{w}g'_e) = 0 \\ d_2\sqrt{u}(a_e + \sqrt{w}a'_e) + b_2v_2(c_e + \sqrt{w}c'_e)) + c_2v_3(u_e + \sqrt{w}u'_e) + a_2v_1(g_e + \sqrt{w}g'_e) = 0 \\ d_3\sqrt{u}(a_e + \sqrt{w}a'_e) + b_3v_2(c_e + \sqrt{w}c'_e)) + c_3v_3(u_e + \sqrt{w}u'_e) + a_3v_1(g_e + \sqrt{w}g'_e) = 0 \end{cases}$$

Using similar method with proof of Property 1, we can get

$a_e = g_e = c_e = u_e = a'_e = g'_e = c'_e = u'_e = 0$. Therefore, $e = 0$, which means no circuit exists in this graphical representation.

Property 4 The 3D representation possesses reflection symmetry.

Proof:Usually the sequence is expressed in the order from $5'$ to $3'$. Suppose that the 3D representation for RNA secondary structure is described by $(x_i, y_i, z_i), i = 0, 1, 2, \ldots, N$. Suppose again that the 3D representation for the reverse structure, i.e, the same sequence but from $3'$ to $5'$ is described by $(\hat{x}_i, \hat{y}_i, \hat{z}_i)$, we find

$$\begin{cases} \hat{x}_i = x_N - x_{N-i}, \\ \hat{y}_i = y_N - y_{N-i}, \\ \hat{z}_i = z_N - z_{N-i}. \end{cases} \tag{4}$$

4 Numerical characterizations

In this section, we give a numerical characterization of the new representation that will facilitate quantitative comparisons of RNA secondary structure. One of the possibilities to achieve this aim is to characterize the curves by invariants. In order to find some of the invariants sensitive to the form of the curve, one can transform the graphical representation of the curve into another mathematical object, a matrix. Once a matrix representation of a RNA sequence is given, some of matrix invariants, e.g. the leading eigenvalues, can be used as descriptors of the sequence [3]. Here, we consider the quotient matrix E/P and E/G[3]. The (i,j) element of matrix E/P is defined to be the quotient of the Euclidean-distance between vertices i and j of the 3DR-Curve and the sum of the distances between the same pair of vertices. In other words, $[E/P]_{ij} = [ED]_{ij} / \sum_{k=i}^{j-1} [ED]_{k,k+1}$, where $[ED]_{ij}$ is the Euclidean distance between a pair of vertices and the The (i,j) element $[E/G]_{ij}$ of matrix E/G is defined to be $[ED]_{ij}/|i-j|$.

We choose the leading eigenvalues of Quotient Matrices E/P and E/G as descriptors of RNA sequences. Since the 3DR-Curve does not represent the genuine molecular geometry we are not interested in the interpretation of the leading eigenvalues of these matrices, but are interested in them as numerical parameters that may facilitate comparisons of RNA sequences.

5 Similarities/Dissimilarities

We will illustrate the use of the 3DR-Curve of RNA secondary structure with the examination of similarities/dissimilarities among the 5 RNA pseudoknots in Figure 3 and Figure 4, which were reported in [15]. We can obtain different information about the similar RNA sequence if we choose different parameters of the 3DR-Curves.

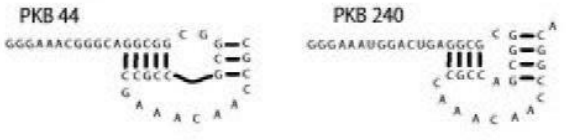

Figure 3: pseudoknots

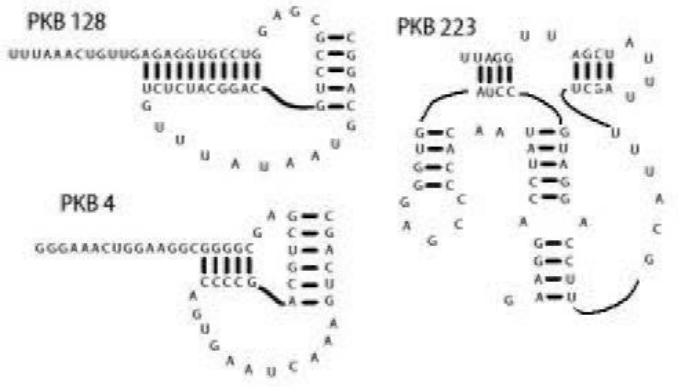

Figure 4: pseudoknots

In Fig. 5, we show the 3DR-Curves of the 5 RNA pseudoknots in Figure 3 and Figure 4 under $v1 = 7; v2 = 1; v3 = 8; u = 2; d = 2; w = 3; \alpha = -\pi/5; \beta = \pi/2; \gamma = \pi/4; \theta = \pi/3;$. By examining these 3DR-Curves, we find that $PKB128$ and $PKB223$ are dissimilar to others, and the similar species should be $PKB4$, $PKB44$ and $PKB240$. But we can clearly found which one is more similar with $PKB240$, or with $PKB44$. So we need change the parameters so that we can analyze these RNA sequences by corresponding different forms of 3DR-Curve. For example, 3DR-Curves of the 5 RNA pseudoknots with $v1 = -2; v2 = -4; v3 = 3; u = 9; d = 2; w = 3; \alpha = \pi/3; \beta = \pi/2; \gamma = \pi/4; \theta = \pi/6$, respectively, are drawing in Fig. 6.

Observing figure 6, we can see the curve of $PKB128$ have some similar profile and tendency with $PKB223$. So we think $PKB128$ is more similar with $PKB223$. We also can find that $PKB240$ has various degree of leaps comparing with $PKB44$, the tendency of

3DR-Curve of $PKB4$ is different from that of $PKB240$. That is $PKB240$ is more similar with $PKB44$ than $PKB4$.

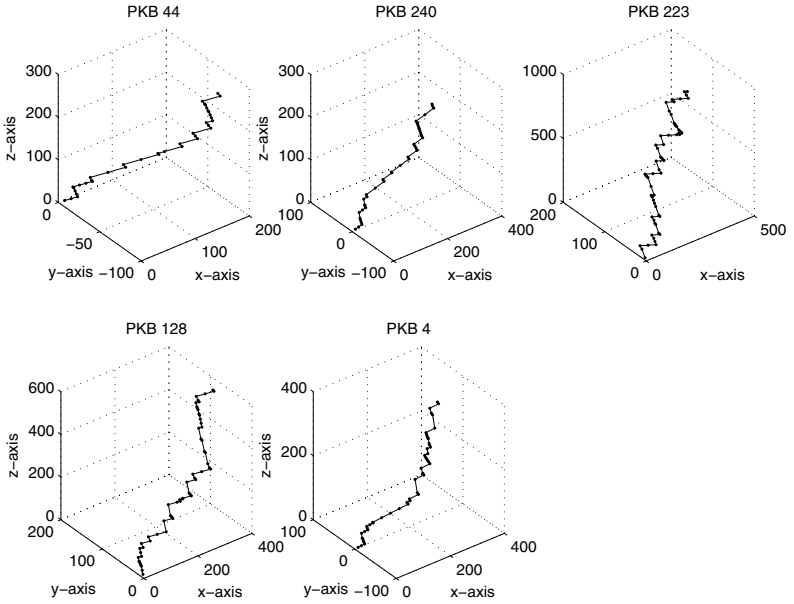

Figure 5: pseudoknots

We can conclude that different pattern can show us different information about the RNA sequences. Of course, it is necessary to analyze the similarity by other numerical characterizations of 3DR-Curves.

A direct comparison of RNA pseudoknots using computer codes is somewhat less straightforward due to the fact that the pseudoknots have different lengths and exist in different places. We construct a 4-component vector consisting of the normalized leading eigenvalue of the Quotient Matrix E/G or E/P of the 3DR-Curves with different parameters. For our application, we will use the following four 3DR-Curves:

Case a. 3DR-Curve corresponding to $v1 = 2; v2 = 4; v3 = 3; u = 9; d = 2; w = 3; \alpha = \pi/3; \beta = \pi/2; \gamma = \pi/4; \theta = \pi/6$ in Eq.(3).

Case b. 3DR-Curve corresponding to $v1 = 7; v2 = 1; v3 = 8; u = 2; d = 2; w = 3; \alpha = \pi/5; \beta = \pi/2; \gamma = \pi/4; \theta = \pi/3$ in Eq.(3).

Case c. 3DR-Curve corresponding to $v1 = 1; v2 = 3; v3 = 4; u = 5; d = 3; w = 3; \alpha = \pi/6; \beta = \pi/2; \gamma = \pi/9; \theta = \pi/3$ in Eq.(3).

5 Similarities/Dissimilarities

We will illustrate the use of the 3DR-Curve of RNA secondary structure with the examination of similarities/dissimilarities among the 5 RNA pseudoknots in Figure 3 and Figure 4, which were reported in [15]. We can obtain different information about the similar RNA sequence if we choose different parameters of the 3DR-Curves.

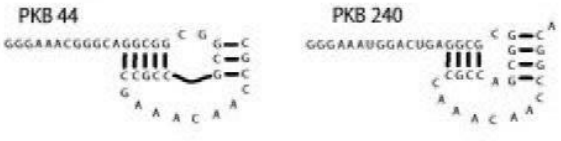

Figure 3: pseudoknots

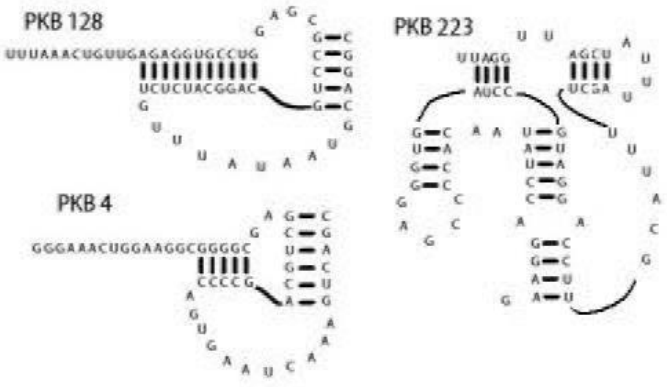

Figure 4: pseudoknots

In Fig. 5, we show the 3DR-Curves of the 5 RNA pseudoknots in Figure 3 and Figure 4 under $v1 = 7; v2 = 1; v3 = 8; u = 2; d = 2; w = 3; \alpha = -\pi/5; \beta = \pi/2; \gamma = \pi/4; \theta = \pi/3;$. By examining these 3DR-Curves, we find that $PKB128$ and $PKB223$ are dissimilar to others, and the similar species should be $PKB4$, $PKB44$ and $PKB240$. But we can clearly found which one is more similar with $PKB240$, or with $PKB44$. So we need change the parameters so that we can analyze these RNA sequences by corresponding different forms of 3DR-Curve. For example, 3DR-Curves of the 5 RNA pseudoknots with $v1 = -2; v2 = -4; v3 = 3; u = 9; d = 2; w = 3; \alpha = \pi/3; \beta = \pi/2; \gamma = \pi/4; \theta = \pi/6$, respectively, are drawing in Fig. 6.

Observing figure 6, we can see the curve of $PKB128$ have some similar profile and tendency with $PKB223$. So we think $PKB128$ is more similar with $PKB223$. We also can find that $PKB240$ has various degree of leaps comparing with $PKB44$, the tendency of

Table 3: The leading normalized eigenvalues of the E/P

EP	$PKB44$	$PKB240$	$PKB223$	$PKB128$	$PKB4$
a	0.9431	0.9410	0.9462	0.9480	0.9436
b	0.9556	0.9591	0.9723	0.9714	0.9651
c	0.9399	0.9418	0.9392	0.9358	0.9414
d	0.9528	0.9514	0.9402	0.9398	0.9494

In table 4, 5, The similarity/dissimilarity matrix for the 5 peseudoknots based on the Euclidean distances between the end points of the 4-component vectors of the leading normalized eigenvalues of the E/G and E/P matrices

Observing Table 4, 5, we find the most similar species pairs should be ($PKB240$, $PKB44$),($PKB128$, $PKB223$),($PKB44$, $PKB4$) and ($PKB240$, $PKB4$). But the dissimilar species pairs are not clear.

Table 4: The similarity/dissimilarity matrix for the 5 peseudoknots based on E/G

EG	$PKB44$	$PKB240$	$PKB223$	$PKB128$	$PKB4$
$PKB44$	0	0.7616	5.8363	4.8762	1.6802
$PKB240$		0	5.7948	4.9027	1.3808
$PKB223$			0	1.3432	4.4430
$PKB128$				0	3.6328
$PKB4$					0

Table 5: The similarity/dissimilarity matrix for the 5 peseudoknots based on E/P

EP	$PKB44$	$PKB240$	$PKB223$	$PKB128$	$PKB4$
$PKB44$	0	0.0047	0.0211	0.0214	0.0102
$PKB240$		0	0.0182	0.0192	0.0068
$PKB223$			0	0.0040	0.0121
$PKB128$				0	0.0135
$PKB4$					0

6 Conclusion

High complexity and degeneracy are major problems in previous RNA secondary structure representations. Our representation provides a direct plotting method to denote RNA secondary structures without degeneracy. From the RNA graph, the A,U,G,C,A-U and C-G usage as well as the original RNA structure can be recaptured mathematically without loss of

textual information. The current 3-dimensional graphical representation of RNA secondary structures provides different approaches for both computational scientists and molecular biologists to analysis RNA secondary structures efficiently.

References

[1] Bo Liao, Tianming Wang, Analysis of similarity of DNA sequences based on 3D graphical representation, *Chem. Phys. Lett.*, 388(2004), 195-200.

[2] M. Randic, M. Vracko, A. Nandy, S. C. Basak, On 3-D graphical representation of DNA primary sequence and their numerical characterization, *J. Chem. Inf. Comput. Sci.*, 40(2000), 1235-1244.

[3] Milan Randic, Majan Vracko, Nella Lers, Dejan Plavsic, Novel 2-D graphical representation of DNA sequences and their numberical characterization, *Chem. Phys. Lett.*, 368(2003), 1-6.

[4] Hamori,E., Novel DNA sequence representations, Nature, 314(1985), 585-586.

[5] Gates, M. A., Simple DNA sequence representations, Nature, 316(1985), 219-223.

[6] A. Nandy, A new graphical representation and analysis of DNA sequence structure: I. Methodology and Application to Globin Genes, *Curr. Sci.*, 66(1994), 309-314.

[7] Nandy, A. Two-dimensional graphical representation of DNA sequences and intron-exon discrimination in intron-rich sequences, Comput. Appl. Biosci., 12(1996),55-62.

[8] Bo Liao, A 3D graphical representation of DNA sequence, *Chem. Phys. Lett.*, 401(2005), 196-199.

[9] Bo Liao, Yusen Zhang, Kequan Ding, Tianming Wang, Analysis of similarity/dissimilarity of DNA sequences based on a condensed curve representation,*J. Mol. Struct.(Theochem)*, 717(2005), 199-203.

[10] Gan, H. H., Pasquali, S., Schlick, T., Exploring the repertoire of RNA secondary motifs using graph theory:implications for RNA design.*Nuclei Acids Res.*,31(2003), 2926-2943.

[11] Shapiro, B. A., Zhang, K. Z., Comparing multiple RNA secondary structure using tree comparisons.*Comput. Biomed. Res.*, 6(1990), 309-318.

[12] Bo Liao, Tianming Wang, A 3D graphical representation of RNA secondary structure,*J. Biomol. Struc. Dynamics*, 21(2004), 827-832.

[13] Bo Liao, Kequan Ding,Tianming Wang, On a six-dimensional representation of RNA secondary structures, *J. Biomol. Struc. Dynamics* , 22 (2005), 455-464.

[14] Bo Liao, Tianming Wang, Kequan Ding, On a seven-dimensional representation of RNA secondary structures, *Lecture Series on Compute and Computationa Science*,1(2004), 310-312.

[15] Samuela Pasquali, Hin Hark Gan and Tamar Schlick, Modular RNA architecture revealed by computational analysis of existing pseudoknots and ribosomal RNAs, Nucleic Acids Res., 33(2005), 1384-1398.