MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **57** (2007) 697-710

ISSN 0340 - 6253

On 2D Graphical Representation of RNA Secondary Structure

Yusen Zhang*
Department of Applied Mathematics, Shandong University at Weihai
Weihai 264209, China

(Received October 11, 2006)

Abstract. A 2D graphical representation of RNA secondary structures has been derived for mathematical denotation of RNA structure. The two-dimensional graphical representation avoids loss of information accompanying alternative 2D and 3D representation in which the curve standing for RNA structure overlaps and intersects itself. And we use the 2DR-Curve of RNA secondary structure with the examination of similarities/dissimilarities among nine species based on invariant *inv* and eigenvalue of the quotient matrices of the 2DR-Curve.

1 Introduction

In recent years several authors outlined different approaches to compute the similarity of DNA sequences based on 2-D or 3-D graphical representation[1-8]. Graphical representation of DNA sequence provides a simple way of viewing, sorting and comparing various gene structures.

Ribonucleic acid(RNA) is an important molecule which performs a wide range of functions in the biological system. In particular, it is RNA(not DNA) that contains genetic information of virus such as HIV and therefore regulates the functions of such virus. RNA has recently become the center of much attention because of its catalytic properties, leading to an increased interest in obtaining structural information. Current RNA secondary structure comparison algorithms have focused exclusively on tree structures owing to their relative simplicity for quantitative analysis[9-11]. But tree structures refer to mathematical constructs for RNA secondary structures without psedoknots. So we should present a new representation to analyze and to compare RNA secondary structures with psedoknots. Similar with the graphical representations of DNA sequences, we also can outline several graphical representations of RNA primary sequences based on 2D and 3D to compute the similarity of RNA primary sequences. Recently, Liao have proposed 3D and 6D graphical representation of RNA secondary structures[12,13], but the representation is not unique.

*Corresponding author:zhangys@sdu.edu.cn

Here, we present a two-dimensional graphical representation of RNA secondary structures, which avoids the limitation associated with non-crossing. The RNA psedoknots can also be represented as two-dimensional graphical representations. The similarities and dissimilarities among nine species based on invariant *inv* and eigenvalue of the quotient matrices of the 2DR-Curve also be considered in this paper.

2 2D 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 and G-C. Let A', U', G', C' denote A, U, G, C in the base pairs A-U and G-C, 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 $C'U'G'G'C'G'AUUGCG'A'G'A'C'C'A'UGU$ $C'G'C'C'A'G'CUCU'G'G'U'C'U'C'CA$ (from $3'$ to $5'$)(see Figure 1).

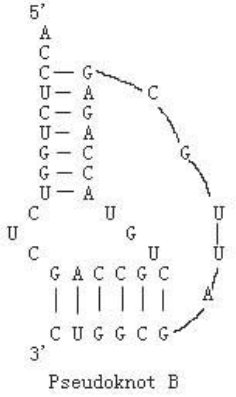


Fig. 1: pseudoknot

We will illustrate the two-dimensional characterization of RNA secondary structure. We construct a map between the bases of characteristic sequences and plots in 2D space, then we will obtain a 2D representation of the corresponding RNA secondary structure. In 2D space points, vectors and directions have four components, and we will assign the following basic elementary directions to the four free bases and three base pairs.

$(1,0) \longrightarrow A$, $(0,\sqrt{k}) \longrightarrow C$, $(\sqrt{m},0) \longrightarrow A'$, $(0,\sqrt{km}) \longrightarrow C'$,
$(0,1) \longrightarrow G$, $(\sqrt{k},0) \longrightarrow U$, $(\sqrt{km},0) \longrightarrow G'$, $(0,\sqrt{m}) \longrightarrow U'$,

where k, m, km are different positive real numbers but not perfect square numbers.

So that we will reduce a RNA secondary structure into 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

$$\begin{cases} x_i = a_i + \sqrt{k}u_i + \sqrt{m}a'_i + \sqrt{km}g'_i, \\ y_i = g_i + \sqrt{k}c_i + \sqrt{km}c'_i + \sqrt{m}u'_i. \end{cases} \tag{1}$$

where $a_i, c_i, g_i, u_i, a'_i, c'_i, g'_i$ and u'_i are the cumulative occurrence numbers of A, C, G , U, A', C', G', and U', respectively, in the subsequence from the 1st base to the i-th base in the sequence. We define $a_0 = c_0 = g_0 = u_0 = a'_0 = c'_0 = g'_0 = u'_0 = 0$.

We call the corresponding plot set be characteristic plot set. The curve connecting all plots of the characteristic plot set in turn is called 2DR-Curve (2D Curve of RNA), which is determined by a, k, m that satisfy above mentioned condition. In Figure 2, we show the curve that represent the pseudoknot B(see Figure 1).

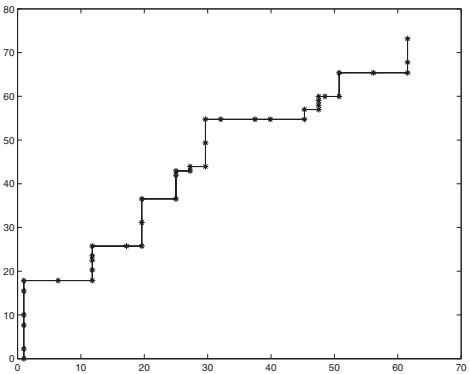


Fig. 2: 2DR-Curves of pseudoknot B with $k = 5; m = 2$

Bases of RNA can be classed into groups, purine(A,G)/pyrimidine(C,U), amino(A,C)/keto(G,U) and week-H bond(A,U)/strong-H band(G,C). We can obtain only three representations corresponding to the three classifications. We call them pattern AG, AC and AU, respectively. The eq.(1) corresponds to 2DR-Curve of RNA secondary structure based on pattern AU.

By assigning the following vectors to the four bases:

$(1,0) \longrightarrow A$, $(0,\sqrt{k}) \longrightarrow G$, $(\sqrt{m},0) \longrightarrow A'$, $(0,\sqrt{km}) \longrightarrow G'$,
$(0,1) \longrightarrow C$, $(\sqrt{k},0) \longrightarrow U$, $(\sqrt{km},0) \longrightarrow C'$, $(0,\sqrt{m}) \longrightarrow U'$,

then, we get 2DR-Curve of RNA secondary structure based on pattern AC.

$$\begin{cases} x_i = a_i + \sqrt{k}u_i + \sqrt{m}a'_i + \sqrt{km}c'_i, \\ y_i = c_i + \sqrt{k}g_i + \sqrt{km}g'_i + \sqrt{m}u'_i. \end{cases} \tag{2}$$

Assigning the following vectors to the four bases:

$(1,0) \longrightarrow A$, $(0,\sqrt{k}) \longrightarrow G$, $(\sqrt{m},0) \longrightarrow A'$, $(0,\sqrt{km}) \longrightarrow G'$,
$(0,1) \longrightarrow U$, $(\sqrt{k},0) \longrightarrow C$, $(\sqrt{km},0) \longrightarrow U'$, $(0,\sqrt{m}) \longrightarrow C'$,

then, we get 2DR-Curve of RNA secondary structure based on pattern AU.

$$\begin{cases} x_i = a_i + \sqrt{k}c_i + \sqrt{m}a'_i + \sqrt{km}u'_i, \\ y_i = u_i + \sqrt{k}g_i + \sqrt{km}g'_i + \sqrt{m}c'_i. \end{cases} \tag{3}$$

3 Properties

Property 1 For a given RNA secondary structure there is a unique 2DR-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

$$\begin{cases} x_i = a_i + \sqrt{k}u_i + \sqrt{m}a'_i + \sqrt{km}g'_i, \\ y_i = g_i + \sqrt{k}c_i + \sqrt{km}c'_i + \sqrt{m}u'_i. \end{cases}$$

If x_i, y_i can also be expressed as

$$\begin{cases} x_i = A_i + \sqrt{k}U_i + \sqrt{m}A'_i + \sqrt{km}G'_i, \\ y_i = G_i + \sqrt{k}C_i + \sqrt{km}C'_i + \sqrt{m}U'_i. \end{cases}$$

then, we have

$(a_i - A_i) + \sqrt{k}(u_i - U_i) + \sqrt{m}(a'_i - A'_i) + \sqrt{km}(g'_i - G'_i) = 0,$

$(g_i - G_i) + \sqrt{k}(c_i - C_i) + \sqrt{km}(c'_i - C'_i) + \sqrt{m}(u'_i - U'_i) = 0.$

then, we get $a_i = A_i, u_i = U_i, g_i = G_i, c_i = C_i, a'_i = A'_i, u'_i = U'_i, g'_i = G'_i, c'_i = C'_i$.

So, for given x-projection and y-projection of any point $P = (x, y)$ on the structure, 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 can be found by solving linear system Eq.(2). By successive x-projection and y-projection of points on the sequence, we can recover the original RNA secondary structure uniquely from the 2DR-Curve.

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 at most has 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 $\Delta x_i = a$ and $\Delta y_i = w_2 a$. This result is independent of the conformation state of the (i-1)-th residue. The eight numbers ($\Delta x_i = a, \Delta y_i = w_2 a$) 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).

Property 3 There is no circuit or degeneracy in our 2DR-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:

$a_e + \sqrt{k}u_e + \sqrt{m}a'_e + \sqrt{km}g'_e = 0$

$g_e + \sqrt{k}c_e + \sqrt{km}c'_e + \sqrt{m}u'_e = 0$

Clearly Eq.(3) hold if , and only if $a_e = g_e = c_e = u_e = a'_e = g'_e = c'_e = u'_e = 0$. Therefore, $e = 0$, which means no circuit exists in this graphical representation.

Table 1: Eight possible direction

	Δx_i	Δy_i	$\lvert\Delta r_i\rvert$
A	1	0	1
G	0	1	1
C	0	$\sqrt{k}$	$\sqrt{k}$
U	$\sqrt{k}$	0	$\sqrt{k}$
A'	$\sqrt{m}$	0	$\sqrt{m}$
G'	$\sqrt{km}$	0	$\sqrt{km}$
C'	0	$\sqrt{km}$	$\sqrt{km}$
U'	0	$\sqrt{m}$	$\sqrt{m}$

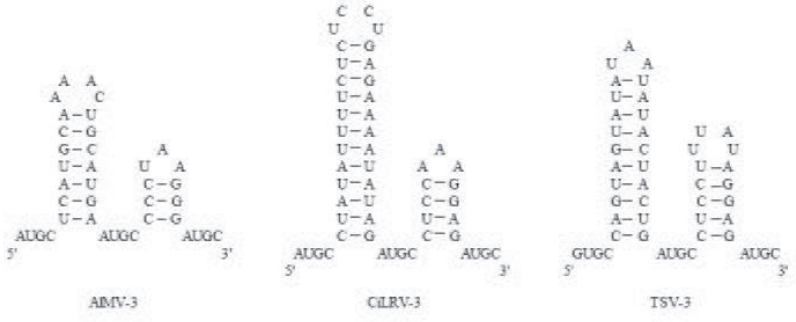
Property 4 The 2D representation possesses the reflection symmetry.

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

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

4 Visualization of RNA secondary structure

In this section we will make a comparison for the secondary structures at the 30-terminus belonging to nine different species based on 3DR-Curves. In Figure 3, the secondary structures at the 30-terminus belonging to nine different viruses are listed, which were reported by Reusken and Bol[16].



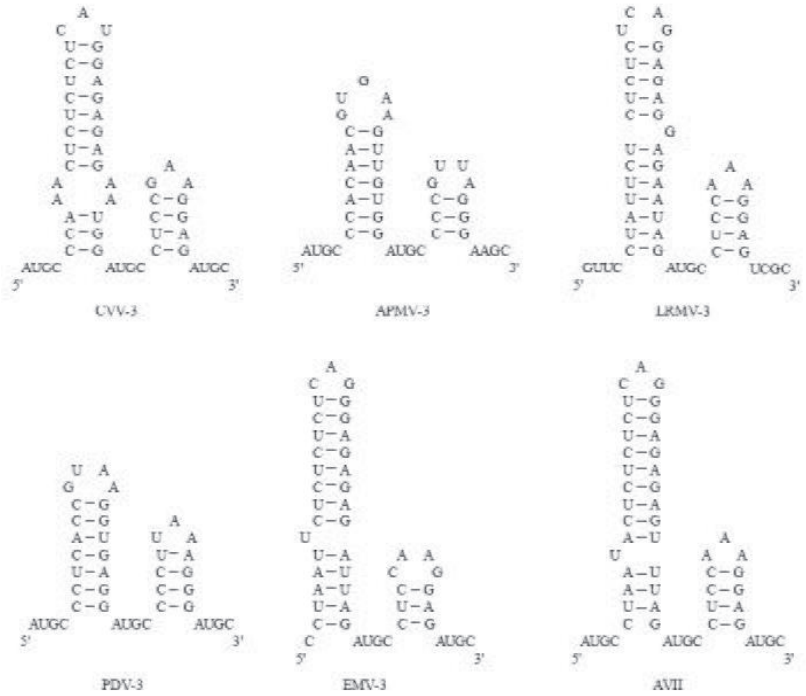


Fig. 3: Secondary structure at the 30-terminus of RNA 3 of alfalfa mosaic virus (AlMV-3), citrus leaf rugose virus (CiLRV-3), tobacco streak virus(TSV-3), citrus variegation virus (CVV-3), apple mosaic virus (APMV-3), prune dwarf ilarvirus (PDV-3), elm mottle virus (EMV-3) and asparagus virus II (AVII). Numbering of nucleotides is from the 3' end of RNA 3.

In fig. 4, 5, 6, we show the 2DR-Curves for the secondary structures at the 3'-terminus belonging to nine different viruses based on three different pattern (AG, AC and AU). We can find that tendency of one virus along the 2DR-Curve is strictly different from that based on different patterns. By observing the three figures, we also can observe that the 2DR-Curves of LRMV-3, EMV-3 and AVII-3 are very similar to each other, and 2DR-Curve of CVV- 3 is also more similar to the above three 2DR-Curves.

It should be notice that different parameters k, m can result in different visual clues to RNA Secondary Structure. That means the 2DR-Curve can provides visual clues to four bases over different regions of a RNA Secondary Structure and can be examined for local behavior of the RNA Secondary Structure by selecting different k, m.

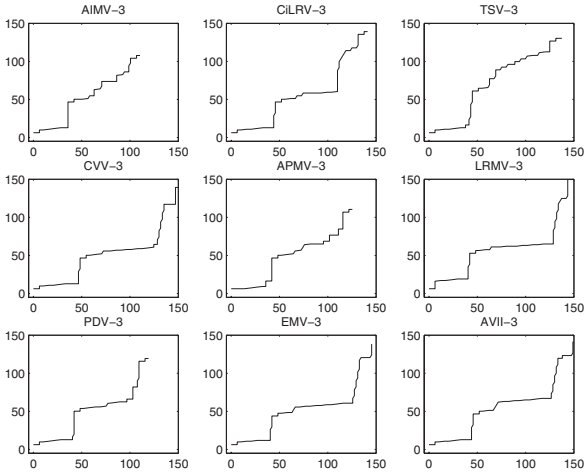


Fig. 4: The 2DR-Curves for the secondary structures at the 3'-terminus belonging to nine different viruses based on AG with k=3,m=13

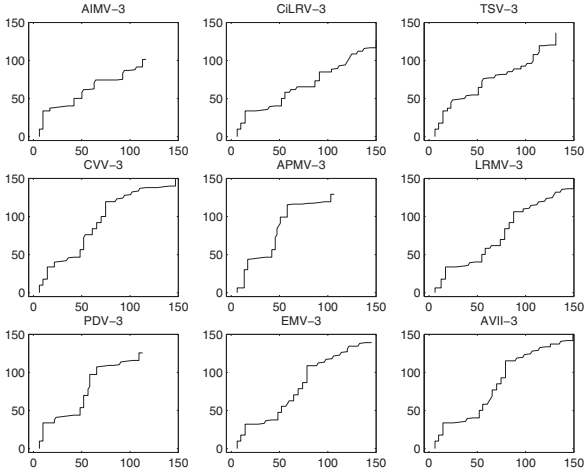


Fig. 5: The 2DR-Curves for the secondary structures at the 3'-terminus belonging to nine different viruses based on AC with k=3,m=13

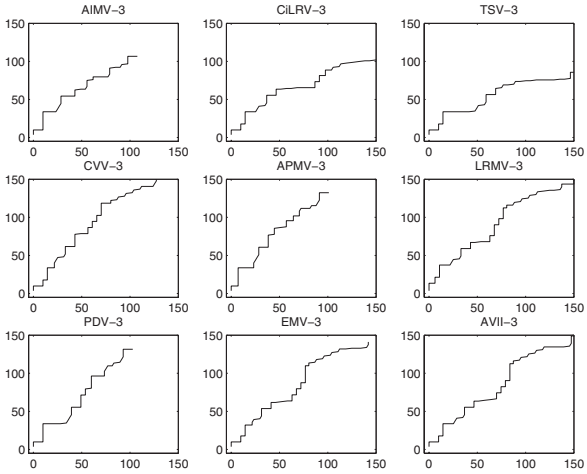


Fig. 6: The 2DR-Curves for the secondary structures at the 3'-terminus belonging to nine different viruses based on AU with k=3,m=13

5 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 [5,6]. Here, we consider the quotient matrix E/P and E/G[7,8]. 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 $[ED]_{ij}/|i-j|$. It is easy to see that quotient matrices E/P and E/G can be regarded as distance-like matrices of 2DR-Curve of DNA sequences.

The leading eigenvalues of quotient matrices E/P and E/G has been interpreted as a measure of the folding degree of chain molecules. We choose the leading eigenvalues of quotient matrices E/P and E/G as descriptors of RNA sequences in following discussion.

In paper [14], the authors proposed a new sequence invariant Inv based on 2DD-curves which is defined by

$$Inv(M) = \frac{1}{n-1}\sum_{i=1}^{n}(\sum_{j=1}^{n} a_{ij}) \tag{5}$$

where M is one quotient matrix of 2DD-Curve of DNA sequences.

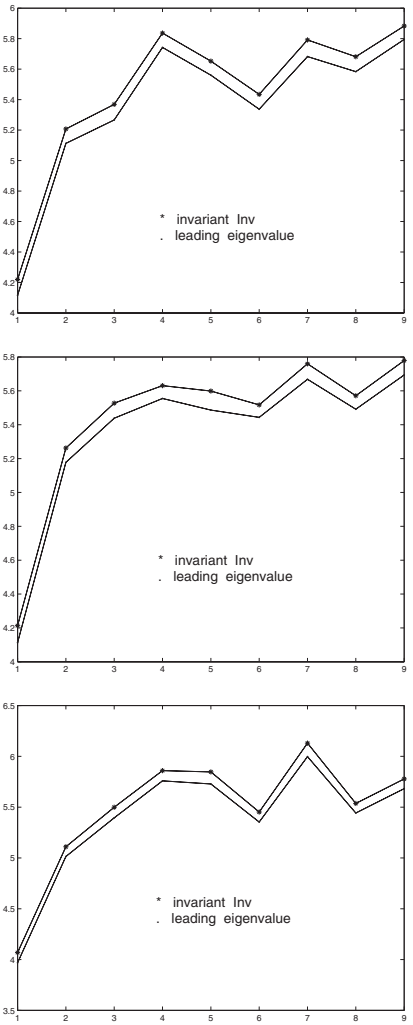


Fig. 7: The invariants Inv VS leading eigenvalues of matrices E/G

Since quotient matrices E/P and E/G can be regarded as distance-like matrices of 2DD-Curve of DNA sequences, Inv is the average value of all the distance-like between any two

points of the 2DD-Curve of a DNA sequence. Since Inv is simple for calculation and thus facilitated for characterization of RNA sequences, we also discuss the application of it based on 2DR-Curve of RNA sequences.

In figure 7, we show the leading eigenvalues and Inv of quotient matrices E/P and E/G of 2DR-Curves for the secondary structures at the 3'-terminus belonging to nine different viruses based on AC, AG, AU, respectively, with $k = 3, m = 13$. We can see the invariant Inv of DNA sequences based on 2DD-Curves also adapt to the matrices E/P and E/G of 2DR-Curves.

It should be notice that the tendency of the curves maybe different if choosing different k, m.

6 Similarities/Dissimilarities

We will illustrate the use of the 2DR-Curve of RNA secondary structure with the examination of similarities/dissimilarities among the the secondary structures at the 3'-terminus belonging to nine different viruses are listed in Fig. 3, which were reported by Reusken and Bol[15]. A direct comparison of these RNA secondary structure using computer codes is somewhat less straightforward due to the fact that the secondary structure have different lengths and exist in different places. We construct a 3-component vector consisting of the normalized leading eigenvalue of the Quotient Matrix E/G or E/P of the 2DR-Curves based on AC, AG and AU, respectively. And use a distance function for vectors to measure the distance between the vectors.

Table 2: The normalized leading eigenvalues of the E/P matrices associated with three different patterns of the 2DR-Curves for the secondary structures at the 3'- terminus belonging to nine viruses (k=3,m=13)

Species	AIMV-3	CiLRV-3	TSV-3	CVV-3	APMV-3	LRMV-3	PDV-3	EMV-3	AVII
AIMV-3	0	0.0557	0.0663	0.0448	0.0289	0.0473	0.0332	0.0583	0.0540
CiLRV-3		0	0.0245	0.0140	0.0322	0.0177	0.0261	0.0185	0.0081
TSV-3			0	0.0350	0.0509	0.0385	0.0406	0.0410	0.0312
CVV-3				0	0.0193	0.0067	0.0143	0.0155	0.0094
APMV-3					0	0.0210	0.0143	0.0304	0.0283
LRMV-3						0	0.0157	0.0120	0.0106
PDV-3							0	0.0276	0.0229
EMV-3								0	0.0110
AVII									0

The analysis of similarity/dissimilarity among RNA secondary structures represented by the 3-component vectors is based on the assumption that two RNA secondary structures are similar if the corresponding 3-component vectors point to a similar direction in the 3D-space and have similar magnitudes. the similarity between these two vectors can be measured

by calculating the Euclidean distance between their end points. Clearly, the smaller is the Euclidean distance the more similar are the two RNA secondary structures.

Table 3: The normalized leading eigenvalues of the E/G matrices associated with three different patterns of the 2DR-Curves for the secondary structures at the 3'- terminus belonging to nine viruses (k=3,m=13)

Species	AIMV-3	CiLRV-3	TSV-3	CVV-3	APMV-3	LRMV-3	PDV-3	EMV-3	AVII
AIMV-3	0	1.1258	1.3784	1.8170	1.4871	1.5242	1.7549	1.6641	1.9376
CiLRV-3		0	0.3435	0.8714	0.4408	0.5073	0.7399	0.6667	0.9378
TSV-3			0	0.7445	0.2001	0.3790	0.4901	0.5884	0.8033
CVV-3				0	0.5457	0.3788	0.3608	0.2359	0.1717
APMV-3					0	0.1951	0.3078	0.4055	0.6080
LRMV-3						0	0.2832	0.2227	0.4573
PDV-3							0	0.3590	0.4060
EMV-3								0	0.2820
AVII									0

In Table 2 and Table 3, we give the similarities and dissimilarities for the secondary structures at the 3'- terminus belonging to nine viruses of Figure 3 based on the Euclidean distances between the end points of the 3-component vectors.

Table 4: The normalized inv/N of the E/P matrices associated with three different patterns of the 2DR-Curves for the secondary structures at the 3'- terminus belonging to nine viruses (k=3,m=13)

Species	AIMV-3	CiLRV-3	TSV-3	CVV-3	APMV-3	LRMV-3	PDV-3	EMV-3	AVII
AIMV-3	0	0.0483	0.0594	0.0375	0.0280	0.0407	0.0323	0.0534	0.0462
CiLRV-3		0	0.0252	0.0149	0.0255	0.0186	0.0199	0.0185	0.0086
TSV-3			0	0.0368	0.0458	0.0404	0.0352	0.0417	0.0327
CVV-3				0	0.0121	0.0067	0.0101	0.0166	0.0090
APMV-3					0	0.0153	0.0145	0.0261	0.0206
LRMV-3						0	0.0127	0.0135	0.0106
PDV-3							0	0.0249	0.0167
EMV-3								0	0.0113
AVII									0

Observing Table 2 and Table 3, we find AIMV-3 is very dissimilar to others among the 9 species because its corresponding row has lager entries. On the other hand, the CCV-3 is very similar to others. In Table 2, we can find the more similar are AVII-3 and CVV-3 with a value of 0.094, LRMV-3 and CVV-3 with a value 0.0067, AVII-3 and CiLRV-3 with a value of 0.0081 and AVII-3 and LRMV-3 with a value of 0.0106. In Table 3, the similarities and dissimilarities for 9 species that based on the normalized leading eigenvalues of the E/G

matrices. Observing Table 3, we find the more similar species pairs are APMV-3-LRMV-3, AVII-3-CVV-3, TSV-3-APMV-3,EMV-3-LRMV-3 and CVV-3-EMV-3.

We will characterize the coding sequences of the RNA secondary structure of nine species, shown in Figure 3, by means of the invariant *inv* of the E/G and E/P matrices. We also construct the 3-component vector consisting of the normalized inv/N, where N is the number of bases making up the corresponding RNA sequence, of the Quotient Matrix E/G or E/P of the 2DR-Curves based on AC, AG and AU, respectively. the normalized inv/N of the E/G and E/P matrices associated with three different patterns of the 2DR-Curves with $k = 3, m = 13$ for the secondary structures at the 3'- terminus belonging to nine viruses.

Table 5: The normalized inv/N of the E/G matrices associated with three different patterns of the 2DR-Curves for the secondary structures at the 3'- terminus belonging to nine viruses (k=3,m=13)

Species	AIMV-3	CiLRV-3	TSV-3	CVV-3	APMV-3	LRMV-3	PDV-3	EMV-3	AVII
AIMV-3	0	1.1030	1.3630	1.7829	1.4943	1.4910	1.7508	1.6328	1.9062
CiLRV-3		0	0.3483	0.8701	0.4728	0.5067	0.7688	0.6649	0.9331
TSV-3			0	0.7361	0.2113	0.3769	0.5132	0.5823	0.7881
CVV-3				0	0.5324	0.3757	0.3452	0.2405	0.1788
APMV-3					0	0.2056	0.3101	0.4037	0.5838
LRMV-3						0	0.3091	0.2225	0.4533
PDV-3							0	0.3683	0.3820
EMV-3								0	0.0110
AVII									0

In Table 4 and Table 5, we have listed the similarity and dissimilarity tables between the secondary structures of nine species based on invariant *inv*. The Euclidean distance between 3-component vectors consisting of the normalized leading eigenvalue and 3-component vectors consisting of the normalized inv/N is different measures of the similarity of RNA secondary structures. Observing Table 4 and Table 5, we can obtain the similar results with that in Table 2 and Table 3. As one can see there exists an overall qualitative agreement among similarities based on different descriptors despite some variations among them. However, the *Inv* is more simple for calculation than the eigenvalues of the matrices and thus facilitated for characterization of RNA secondary structures

7 Conclusion

We have proposed a 2D representation based on the classifications of bases and base pairs, and presented two similarity measure between RNA secondary structures. A simple 2D representation substitutes the complicated molecular structure. Our representation provides a direct plotting method to denote RNA secondary structures which has no circuit. 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 two-dimensional graphical representation of RNA secondary structures provides different approaches for both computational scientists and molecular biologists to analysis RNA secondary structures efficiently with different parameters k and m.

Acknowledgements

This work was supported in part by the Shandong Natural Science Foundation (Y2006A14).

References

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

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

[3] 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.

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

[5] 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.

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

[7] Yusen Zhang, Bo Liao, Kequan Ding, On 2D Graphical Representation of DNA Sequence of Nondegeneracy, *Chem. Phys. Lett.*, 411(2005), 28-32.

[8] Yusen Zhang, Bo Liao, Kequan Ding, On 3DD-Curves of DNA Sequences, *Molecular Simulation*, 32(2006), 29-34.

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

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

[11] S. Y. Le, R. Nussinov, J. V. Maizel, Tree graphs of RNA secondary structures and their comparisons,*Comput. Biomed.Res.*,22(1989), 461-473.

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

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

[14] Yusen Zhang, Wei Chen, Invariants of DNA Sequences Based on 2DD-Curves, *J. Theor. Biol.*, 242 (2006), 382-388.

[15] B. E. Chantal, M. Reusken, F. Bol. John, Structural elements of the 3'-terminal coat protein binding site in alfalfa mosaic virus RNAs, *Nucl. Acids Res.* 14 (1996), 2660-2665.