Available online at www.sciencedirect.com

Journal of Molecular Graphics and Modelling 25 (2006) 340–344

Journal of Molecular Graphics and Modelling

www.elsevier.com/locate/JMGM

A novel 2D graphical representation of DNA sequences and its application

Qi Dai [a,*], Xiaoqing Liu [a], Tianming Wang [a,b]

[a] Department of Applied Mathematics, Dalian University of Technology, Dalian 116024, China
[b] Department of Mathematics, Hainan Normal University, Haikou 571158, China

Received 24 October 2005; received in revised form 13 December 2005; accepted 15 December 2005
Available online 3 March 2006

Abstract

In this paper, we introduce a novel 2D graphical representation of DNA sequences, the W-curve, which is embedded in two unit circles. We associate the W-curves with the classifications of the nucleotides according to their chemical properties. Then we obtain an 8-component vector with entries being the average sums of the abscissa and *y*-axis of A+C and T+G (A+T and C+G, A+G and T+C), respectively. The introduced vector results in simpler characterizations and comparisons of DNA sequences. The construction of the W-curve has some important advantages: (1) it avoids loss of information and the W-curve standing for DNA doesn't overlap or intersect itself; (2) the space the W-curve occupied is very small, just two unit circles. The utility of the approach can be illustrated by the examination of similarities/dissimilarities among the coding sequences of the first exon of β-globin gene of eleven different species in Table 1.

Keywords: DNA; Graphical representation; Similarity analysis

1. Introduction

The number of DNA sequences as strings of four nucleotides: A (adenine), C (cytosine), G (guanine), T (thymine) is growing rapidly in the DNA database. But it is difficult to obtain information from DNA sequences directly. Therefore, many kinds of methods have been proposed to characterize DNA sequences.

Some researchers introduced an alternative way to compare DNA sequences, based on a set of invariants of DNA sequences, rather than directly using string comparison. Graphical representation of DNA sequences provides a simple way of viewing, sorting, and comparing various gene structures [1–15]. Nandy [8] presented a graphical representation by assigning A (adenine), G (guanine), T (thymine), and C (cytosine) to four direction $(-x)$, $(+x)$, $(-y)$, $(+y)$, respectively. Such a representation of DNA is accompanied with (1) some loss of visual information associated with crossing and overlapping of the curve with itself; (2) an arbitrary decision with respect to the choice of the direction for the bases. Recently, several authors have outlined different graphical representations of the DNA sequences based on 2D, 3D or 4D [3–6,10–15], but some representations [10–12] are also accompanied by some loss of information due to overlapping and crossing of the curve with itself. Then they constructed the D/D, L/L and high order matrices to extract the leading eigenvalues as invariants to characterize the DNA sequences, so their computation is very complex. Furthermore, their representations need a large memory if the sequence is long.

Here, we introduce a novel 2D graphical representation embedded in two unit circles and call it the W-curve. It can avoid the loss of information and uniquely denote the DNA sequence. Finally, we give a simple way to numerically characterize the DNA sequences.

2. 2D graphical representation of DNA sequences and its properties

From the knowledge of biology, we know that the four DNA bases can be classified as $R = \{A, G\}$ and $Y = \{C, T\}$, $M = \{A, C\}$ and $K = \{G, T\}$, $W = \{A, T\}$ and $S = \{G, C\}$ according to their chemical properties. Based on the classification of the four nucleotides, we construct two maps φ_1, φ_2 between the

* Corresponding author.
E-mail addresses: lxq785348@163.com (Q. Dai), wangtm@dlut.edu.cn (T. Wang).

1093-3263/$ – see front matter © 2005 Elsevier Inc. All rights reserved.
doi:10.1016/j.jmgm.2005.12.004

bases of the DNA sequence and the plots in 2D space. If $G = g_1, g_2, \ldots, g_n$ is the given DNA sequence, we define functions φ_1, φ_2 as following:

$$\varphi_1(g_i) = \begin{cases} \left(\frac{2}{n}(i-1), \sqrt{1-\left(1-\frac{2}{n}(i-1)\right)^2}\right), & \text{if } g_i = \text{A or C}, \\ \left(\frac{2}{n}(i-1), -\sqrt{1-\left(1-\frac{2}{n}(i-1)\right)^2}\right), & \text{if } g_i = \text{T or G}, \end{cases}$$

$$\varphi_2(g_i) = \begin{cases} \left(-\frac{2}{n}(i-1), \sqrt{1-\left(1-\frac{2}{n}(i-1)\right)^2}\right), & \text{if } g_i = \text{A or C}, \\ \left(-\frac{2}{n}(i-1), -\sqrt{1-\left(1-\frac{2}{n}(i-1)\right)^2}\right), & \text{if } g_i = \text{T or G}, \end{cases}$$

where $p_i = \varphi_1(g(i)), q_i = \varphi_2(g(i)) (i = 1, 2, \ldots, n)$. Finally, we connect adjacent points and obtain a curve which can better illustrate the characteristic sequence considered. In Fig. 1, the W-curve is a graphical presentation of the first exon of the human β-globin gene.

Property 1. *For a given DNA sequence, all the dots obtained from the maps*φ_1, φ_2 *are on two unit circles, whose centers are* $(-1, 0)$ *and* $(1, 0)$, *respectively.*

Proof. Given a DNA sequence, the maps φ_1, φ_2 transform the DNA sequence into a series of dots $p_1, p_2, \ldots, p_n, q_1, q_2, \ldots, q_n$ where $p_i = \left(\frac{2}{n}(i-1), \pm\sqrt{1-\left(1-\frac{2}{n}(i-1)\right)^2}\right)$, $q_i = \left(-\frac{2}{n}(i-1), \pm\sqrt{1-\left(1-\frac{2}{n}(i-1)\right)^2}\right)$. So we can calculate

$$\left(\frac{2}{n}(i-1)-1\right)^2 + \left(\pm\sqrt{1-\left(1-\frac{2}{n}(i-1)\right)^2}\right)^2 = 1,$$

$$\left(1-\frac{2}{n}(i-1)\right)^2 + \left(\pm\sqrt{1-\left(1-\frac{2}{n}(i-1)\right)^2}\right)^2 = 1.$$

$$i = 1, 2, \ldots, n$$

So p_i and $q_i (i = 1, 2, \ldots, n)$ are on two unit circles $(x-1)^2 + y^2 = 1$ and $(x+1)^2 + y^2 = 1$, and their centers are $(1, 0)$ and $(-1, 0)$, respectively. □

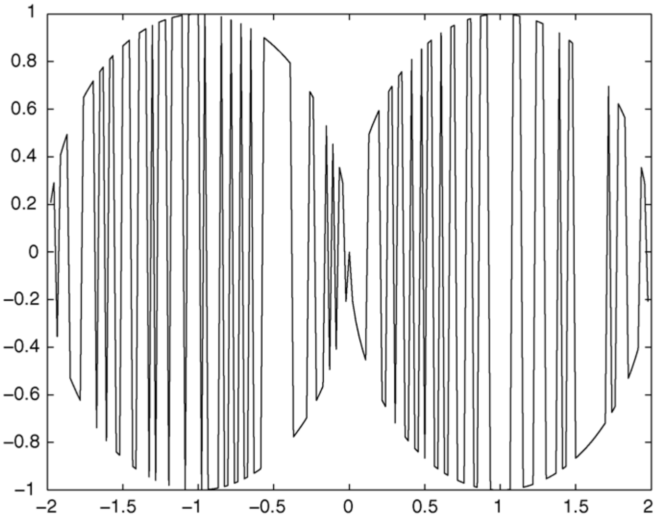

Fig. 1. W-curve of the first exon of human β-globin gene.

Property 2. *For a given sequence, there is a unique 2D representation corresponding to it.*

Proof. From the definition of φ_1, φ_2, we know that a DNA sequence is transformed into a series of dots (p_i, q_i) uniquely. Let $(x_i^1, y_i^1), (x_i^2, y_i^2)$ be the coordinates of p_i, q_i, respectively. If $y_i^1 = \sqrt{1-\left(1-\frac{2}{n}(i-1)\right)^2}$ and $y_i^2 = \sqrt{1-\left(1-\frac{2}{n}(i-1)\right)^2}$, then $g_i = A$; if $y_i^1 = \sqrt{1-\left(1-\frac{2}{n}(i-1)\right)^2}$ and $y_i^2 = -\sqrt{1-\left(1-\frac{2}{n}(i-1)\right)^2}$, then $g_i = C$; if $y_i^1 = -\sqrt{1-\left(1-\frac{2}{n}(i-1)\right)^2}$ and $y_i^2 = -\sqrt{1-\left(1-\frac{2}{n}(i-1)\right)^2}$, then $g_i = T$; if $y_i^1 = -\sqrt{1-\left(1-\frac{2}{n}(i-1)\right)^2}$ and $y_i^2 = \sqrt{1-\left(1-\frac{2}{n}(i-1)\right)^2}$, then $g_i = G$. So every pair of dots (p_i and q_i) corresponds to a nucleotide g_i, and the correspondence is one to one. Therefore, a DNA sequence can be represented uniquely. □

Property 3. *For any* $i = 1, 2, \ldots, n$, *where n is the length of DNA sequence, We denote* $k_{p_i q_i}$ *is the slope of the line determined by two dots* p_i *and* q_i. *If* $k_{p_i q_i}$ *is equal to zero,* g_i *belongs to W; If* $k_{p_i q_i}$ *is not equal to zero,* g_i *belongs to S.*

Proof. Actually

$$k_{p_i q_i} = \frac{\Delta y_i}{\Delta x_i} = \frac{y_i^2 - y_i^1}{x_i^2 - x_i^1}.$$

Δx_i and Δy_i can be calculated from the i th position of the DNA sequence. If the i th residue is A or T, we will get $k_{p_i q_i} = 0$; if the i th residue is G, we find $k_{p_i q_i} < 0$; if the i th residue is C, $k_{p_i q_i} > 0$. So if $k_{p_i q_i}$ is equal to zero, g_i belongs to W; if $k_{p_i q_i}$ is not equal to zero, g_i belongs to S. □

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 the DNA sequence is described by the dots $p_i(x_i^1, y_i^1), q_i(x_i^2, y_i^2) i = 1, 2, \ldots, n$ on the circles. Suppose that the 2D representation for the reverse sequence, i.e., the same sequence but from 3′ to 5′ is described by $p_i^*(x_{1i}^1, y_{1i}^1), q_i^*(x_{2i}^2, y_{2i}^2) \quad i = 1, 2, \ldots, n.$ We obtain

$$\begin{cases} x_{1i}^1 = 2 - x_i^1, \\ y_{1i}^1 = y_i^1. \end{cases}$$

$$\begin{cases} x_{2i}^2 = -2 - x_i^2, \\ y_{2i}^2 = y_i^2. \end{cases}$$

□

3. Application

3.1. The content relation of the nucleotides A, C, G, T of eleven species in Table 1

Given a sequence with the length of n, if we consider the right part of the W-curve, we know the step length is $\frac{2}{n}$ from the definition of the function φ, the abscissa of the first A or C in the sequence must be 0 and the abscissa of the last A or C in the sequence must be $2 - \frac{2}{n}$. Observing Fig. 1, we can find out that every pair of neighboring nucleotides may be both in $\{A, C\}(\{T, G\})$, or if one is in $\{A, C\}$ and the other is in $\{T, G\}$. First we denote $X_i^{\mathrm{A-C}}(X_i^{\mathrm{T-G}})$ the abscissa of A or C (T or G) appearing i th times in the given sequence. Let $X_l^{\mathrm{A-C}}(X_l^{\mathrm{T-G}})$ be the abscissa of the last A or C (T or G). Then we can calculate $\Delta X^{\mathrm{A-C}} = X_i^{\mathrm{A-C}} - X_{i-1}^{\mathrm{A-C}}, i = 2, 3, \ldots, n$. If $\Delta X^{\mathrm{A-C}}$ equals $\frac{2}{n}$, they must be the neighboring nucleotides; if $\Delta X^{\mathrm{A-C}}$ is greater than $\frac{2}{n}$, it is easy to imagine that T or G must appear between $X_{i-1}^{\mathrm{A-C}}$ and $X_i^{\mathrm{A-C}}$, and the number of T and G appearing between $X_{i-1}^{\mathrm{A-C}}$ and $X_i^{\mathrm{A-C}}$ is $\Delta X^{\mathrm{A-C}} \frac{n}{2} - 1$. If the abscissa of the first A or C is zero, A or C must be the first nucleotides; otherwise there must be T or G appearing before the $X_1^{\mathrm{A-C}}$, and the number of T+G appearing before $X_1^{\mathrm{A-C}}$ is $X_1^{\mathrm{A-C}} \frac{n}{2} - 1$. If the abscissa of the last A or C is equal to $2 - \frac{2}{n}$, it means that A or C is the last nucleotides of the considered sequence; Otherwise there must be *T* or G behind the last A or C, and the number of T+G behind the last A or C is $(2 - \frac{2}{n} - x_l^{\mathrm{A-C}}) \frac{n}{2}$. So the total number of T+G is

$$\sum_{i=2}^{l} ((X_i^{\mathrm{A-C}} - X_{i-1}^{\mathrm{A-C}}) \frac{n}{2} - 1) + (2 - \frac{2}{n} - X_l^{\mathrm{A-C}}) \frac{n}{2} + (X_1^{\mathrm{A-C}} \frac{n}{2} - 1).$$

Following the same method, we can compute the numbers of A+C, A+G, A+T, G+C, and C+T, which are listed in Table 2. Taking a closer look at Table 2, we will find the interesting relation, which may give us more information about their evolution. In Table 3, we list the relation of the content of the nucleotides of the sequences in Table 1. Meanwhile we can observe that the longer the $\Delta X^{\mathrm{A-C}}$ is, the more the T and G are contained in this section of the sequence. Therefore, we can find out the longest section of the sequence contains only T or G. Let $\Delta X_m^{\mathrm{A-C}} = \max\{X_i^{\mathrm{A-C}} - X_{i-1}^{\mathrm{A-C}}, 2 - \frac{2}{n} - X_l^{\mathrm{A-C}}, X_1^{\mathrm{A-C}}\}, i = 2, 3, \ldots, n$ denote the longest T and G section.

Observing Table 3, we find most results are similar to the results in Bo Liao et al. [5]. Although there are some difference in Gallus, Opossum, Bovine and Chimpanzee. After validation, our result is correct.

Table 1
The coding sequences of the first exon of β-globin gene of different species

Species	Coding sequence
Human	ATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGCAG
Goat	ATGCTGACTGCTGAGGAGAAGGCTGCCGTCACCGGCTTCTGGGGCAAGGTGAAAGTGGATGAAGTTGGTGCTGAGGCCCTGGGCAG
Opossum	ATGGTGCACTTGACTTCTGAGGAGAAGAACTGCATCACTACCATCTGGTCTAAGGTGCAGGTTGACCAGACTGGTGGTGAGGCCCTTGGCAG
Gallus	ATGGTGCACTGGACTGCTGAGGAGAAGCAGCTCATCACCGGCCTCTGGGGCAAGGTCAATGTGGCCGAATGTGGGGCCGAAGCCCTGGCCAG
Lemur	ATGACTTTGCTGAGTGCTGAGGAGAATGCTCATGTCACCTCTCTGTGGGGCAAGGTGGATGTAGAGAAAGTTGGTGGCGAGGCCTTGGGCAG
Mouse	ATGGTTGCACCTGACTGATGCTGAGAAGTCTGCTGTCTCTTGCCTGTGGGCAAAGGTGAACCCCGATGAAGTTGGTGGTGAGGCCCTGGGCAGG
Rabbit	ATGGTGCATCTGTCCAGTGAGGAGAAGTCTGCGGTCACTGCCCTGTGGGGCAAGGTGAATGTGGAAGAAGTTGGTGGTGAGGCCCTGGGC
Rat	ATGGTGCACCTAACTGATGCTGAGAAGGCTACTGTTAGTGGCCTGTGGGGAAAGGTGAACCCTGATAATGTTGGCGCTGAGGCCCTGGGCAG
Gorilla	ATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGCAGG
Bovine	ATGCTGACTGCTGAGGAGAAGGCTGCCGTCACCGCCTTTTGGGGCAAGGTGAAAGTGGATGAAGTTGGTGGTGAGGCCCTGGGCAG
Chimpanzee	ATGGTGCACCTGACTCCTGAGGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTGGGCAGGTTGGTATCAAGG

Table 2
The content of A+C, A+T, A+G, T+C, C+G, T+G in the given sequences in Table 1

Species	A+C	A+T	A+G	C+T	C+G	T+G
Human	36	37	53	39	55	57
Goat	34	34	52	34	52	52
Opossum	41	43	50	42	49	51
Gallus	43	34	53	39	58	49
Lemur	34	42	54	38	50	58
Mouse	37	40	51	43	54	57
Rabbit	33	37	54	36	53	57
Rat	38	41	53	39	51	54
Gorilla	36	37	54	39	56	57
Bovine	33	35	52	34	51	53
Chimpanzee	40	44	61	44	61	65

Table 3
The relations between a, c, g and t of the 11 species

Species	Relation
Human	g > t > c > a
Goat	g > a = t = c
Opossum	g > t > a > c
Gallus	g > c > a > t
Lemur	g > t > a > c
Mouse	g > t > c > a
Rabbit	g > t > a > c
Rat	g > t > a > c
Gorilla	g > t > c > a
Bovine	g > t > c > a
Chimpanzee	g > t > a = c

3.2. Similarities/dissimilarities among the coding sequences of the first exon of β-globin gene of different species

Comparison of similarities/dissimilarities is the essential motivation of graphical representation, which is reflected in recently published papers [3–6,10–15]. Here we also illustrate the use of our quantitative characterization of the DNA sequences with an examination of similarities/dissimilarities among 11 species in Table 1.

In order to facilitate the quantitative comparison and analysis of different species, we extract some invariants. For a given sequence, the content of the A, C, G and T is fixed, and we use

$$X_1^1 = \frac{\sum_{i=1}^{n}(x_{1i}^{\mathrm{A}} + x_{1i}^{\mathrm{G}})}{n}, \quad X_1^2 = \frac{\sum_{i=1}^{n}(x_{1i}^{\mathrm{C}} + x_{1i}^{\mathrm{T}})}{n},$$

$$Y_1^1 = \frac{\sum_{i=1}^{n}(y_{1i}^{\mathrm{A}} + y_{1i}^{\mathrm{G}})}{n}, \quad Y_1^2 = \frac{\sum_{i=1}^{n}(y_{1i}^{\mathrm{G}} + y_{1i}^{\mathrm{T}})}{n},$$

$$X_2^1 = \frac{\sum_{i=1}^{n}(x_{2i}^{\mathrm{A}} + x_{2i}^{\mathrm{G}})}{n}, \quad X_2^2 = \frac{\sum_{i=1}^{n}(x_{2i}^{\mathrm{C}} + x_{2i}^{\mathrm{T}})}{n},$$

$$Y_2^1 = \frac{\sum_{i=1}^{n}(y_{2i}^{\mathrm{A}} + y_{2i}^{\mathrm{G}})}{n}, \quad Y_2^2 = \frac{\sum_{i=1}^{n}(y_{2i}^{\mathrm{C}} + y_{2i}^{\mathrm{T}})}{n}$$

as invariants, and construct an eight component vector $(X_1^1, X_1^2, X_2^1, X_2^2, Y_1^1, Y_1^2, Y_2^1, Y_2^2)$.

As Bo Liao et al. [5] did, we calculate their correlation angle and the Euclidean distance between the end points of the vectors. The underlying assumption is that if two vectors point in a similar direction in the 2D space, then the two DNA sequences represented by the eight-component vectors are similar. That is to say, the smaller the Euclidean distance between the end-points of two vectors, the more similar the DNA sequences are. Also the smaller the correlation angle between the two vectors, the more similar the DNA sequences are. We define the distance and correlation angle as follow:

Suppose that for two species i and j, the parameters are $X_{i1}^1, X_{i1}^2, Y_{i1}^1, Y_{i1}^2, X_{i2}^1, X_{i2}^2, Y_{i2}^1, Y_{i2}^2$ and $X_{j1}^1, X_{j1}^2, Y_{j1}^1, Y_{j1}^2, X_{j2}^1, X_{j2}^2, Y_{j2}^1, Y_{j2}^2$, respectively. Then we have the distance

$$d_{ij} = \sqrt{\sum_{r=1}^{2}\sum_{t=1}^{2}(X_{ir}^t - X_{jr}^t)^2 + \sum_{r=1}^{2}\sum_{t=1}^{2}(Y_{ir}^t - Y_{jr}^t)^2}.$$

The correlation angle is

$$\theta_{ij} = \arccos \frac{\sum_{r=1}^{2}\sum_{t=1}^{2} X_{ir}^t X_{lr}^t + \sum_{r=1}^{2}\sum_{t=1}^{2} Y_{ir}^t Y_{lr}^t}{\sqrt{\sum_{r=1}^{2}\sum_{t=1}^{2}(X_{ir}^t)^2 + (Y_{ir}^t)^2} \times \sqrt{\sum_{r=1}^{2}\sum_{t=1}^{2}(X_{ir}^t)^2 + (Y_{ir}^t)^2}}$$

In Tables 4 and 5, the data are the similarities and dissimilarities for the 11 coding sequences of Table 1 based on the Euclidean distance between the end points of the eight-

Table 4
The similarity/dissimilarity matrix for the coding sequences of Table 1 based on the Euclidean distances between the end points of eight-component vectors

Species	Human	Goat	Opossum	Gallus	Lemur	Mouse	Rabbit	Rat	Gorilla	Bovine	Chimpanzee
Human	0	0.0169	0.1389	0.1146	0.0603	0.0543	0.0287	0.0704	0.0120	0.0276	0.0155
Goat		0	0.1503	0.1138	0.0700	0.0692	0.0340	0.0849	0.0107	0.0250	0.0266
Opossum			0	0.1221	0.1437	0.0905	0.1569	0.0767	0.1485	0.1581	0.1464
Gallus				0	0.1634	0.1055	0.1432	0.1197	0.1210	0.1367	0.1294
Lemur					0	0.0743	0.0440	0.0702	0.0598	0.0511	0.0515
Mouse						0	0.0719	0.0265	0.0656	0.0769	0.0593
Rabbit							0	0.0827	0.0239	0.0169	0.0159
Rat								0	0.0798	0.0871	0.0738
Gorilla									0	0.0174	0.0172
Bovine										0	0.0236
Chimpanzee											0

Table 5
The similarity/dissimilarity matrix for the coding sequences of Table 1 based on the angle between the eight-component vectors

Species	Human	Goat	Opossum	Gallus	Lemur	Mouse	Rabbit	Rat	Gorilla	Bovine	Chimpanzee
Human	0	0.0540	0.1322	0.1028	0.0658	0.0571	0.0593	0.0820	0.0330	0.0614	0.0301
Goat		0	0.1652	0.1161	0.0628	0.1039	0.0290	0.0719	0.0313	0.0242	0.0353
Opossum			0	0.1140	0.1567	0.0863	0.1711	0.1271	0.1548	0.1749	0.1503
Gallus				0	0.1528	0.1004	0.1391	0.1238	0.1172	0.1371	0.1205
Lemur					0	0.0968	0.0424	0.0629	0.0522	0.0468	0.0460
Mouse						0	0.1067	0.0926	0.0837	0.1110	0.0782
Rabbit							0	0.0695	0.0366	0.0168	0.0316
Rat								0	0.0746	0.0737	0.0691
Gorilla									0	0.0327	0.0170
Bovine										0	0.0359
Chimpanzee											0

component vectors and the correlation angle of the eight-component vectors .

Observing Tables 4 and 5, we find gallus (the only non-mammal among them) and Opossum (the most remote species from the remaining mammals) show larger entries among these species. This is in agreement with the results of Li et al. [15], Liao et al. [5]. Human–Chimpanzee, Human–Mouse, Goat–Bovine, Mouse–Rat, Gorilla–Chimpanzee, Rabbit–chimpanzee have smaller entries, so they are more similar species pairs. This is not an accident, but shows they have close evolutionary relationship.

4. Conclusion

A comparison of DNA sequences even with even fewer than a hundred bases is quite difficult [16]. And as pointed in [14], a direct comparison of the sequences using computer codes is somewhat less straightforward due to the fact that the sequences have different lengths.

In this paper, we have given a novel 2D graphical representation of DNA sequences and similarity analysis of the coding sequences of the first exon of β-globin gene of 11 species. First we outline the novel 2D graphical presentation. Then we give a simple way to extract an 8-component vector as an invariant to characterize the DNA sequence. At last we analyze the similarity and dissimilarity among the eleven different species. Our graphical representation of DNA sequences is visual and direct, without overlapping or intersection, and avoids loss of information.

Acknowledgement

The author thanks all the anonymous referees for their valuable suggestions and support.

It is supported by National Natural Science Foundation of China (10571019).

References

[1] E. Hamori, J. Ruskin, J. Biol. Chem. 258 (1983) 1318.
[2] E. Hamori, BioTechniques 7 (1989) 710.
[3] M. Randić, A.T. Balaban, J. Chem. Inf. Comput. Sci. 43 (2003) 532.
[4] R. Chi, K. Ding, Chem. Phys. Lett. 407 (2005) 63–67.
[5] B. Liao, M. Tan, K. Ding, Chem. Phys. Lett. 402 (2005) 380–383.
[6] B. Liao, T. Wang, J. Mol. Struct.: Theochem. 681 (2004) 209–212.
[7] M.A. Gates, J. Theor. Biol. 119 (1986) 319.
[8] A. Nandy, Curr. Sci. 66 (1994) 309.
[9] A. Nandy, Curr. Sci. 66 (1994) 821.
[10] X. Guo, A. Nandy, Chem. Phys. Lett. 369 (2003) 361.
[11] Y. Wu, A.W. Liew, H. Yan, M.S. Yang, Chem. Phys. Lett. 367 (2003) 170.
[12] A. Nandy, P. Nandy, Chem. Phys. Lett. 368 (2003) 102.
[13] M. Randić, Chem. Phys. Lett. 386 (2004) 468–471.
[14] M. Randić, M. Vračko, N. Lerš, Plavšić, Chem. Phys. Lett. 371 (2003) 202.
[15] C. Li, J. Wang, Comb. Chem. High Throughput Screening 6 (2003) 795–799.
[16] X. Guo, M. Randić, S.C. Bassk, Chem. Phys. Lett. 350 (2001) 106.