

# Spectrum-Like Graphical Representation of DNA Based on Codons

Milan Randić

Laboratory of Chemometrics, National Institute of Chemistry,  
SI – 1001 Ljubljana, Hajdrihova 19, PO Box 660, Slovenia

Received 12-06-2006

## Abstract

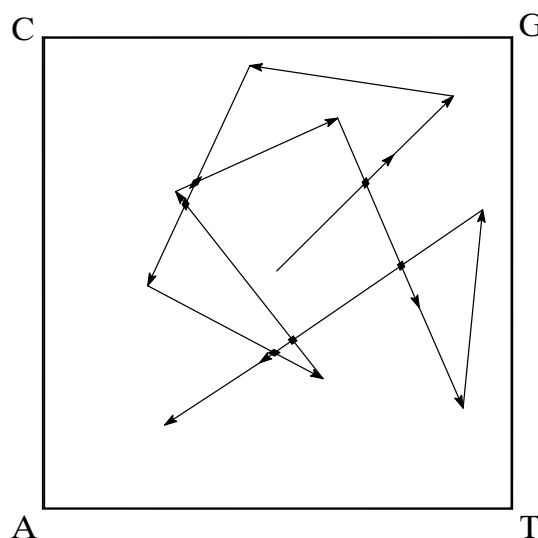
We consider a novel 1-Dimensional representation of DNA, which is based on graphical representations of the 64 triplets of nucleic acids on the periphery of the unit circle. By using the polar coordinate of 64 codons (expressed in radians) four letter DNA sequence is transformed into a numerical sequence with no more than 64 different entries. By depicting the 1-Dim representation (plotted on z-coordinate) and using the x co-ordinate as running index one obtains “spectrum-like” graphical representation of DNA. The novel representation of DNA has some advantages over other spectrum-like 1-dimensional and 2-dimensional representations in using the same coordinates for the same codons thus avoiding computations of coordinates, which is characteristic of the Jeffrey’s algorithm and graphical representations of DNA based on its modification.

**Key words:** graphical representation of DNA,  $\beta$ -globin gene, table of codons, hilbert fractal

## 1. Introduction

Graphical representations of DNA have been initiated about 20 years ago by Hamori and Ruskin,<sup>1,2</sup> and Gates.<sup>3</sup> In the following years they have been expanded by Jeffrey,<sup>4</sup> Pickover,<sup>5</sup> Nandy,<sup>6</sup> and others.<sup>7-9</sup> These graphical representation often start by assuming certain basic geometrical object or format, which is, after adopting a specific algorithm, used for depicting individual nucleic bases. Thus, for instance, Jeffrey based his approach on modifying the algorithm of the mathematical “Chaos game,” designed for construction of various fractal patterns, by selecting a square as the basic geometrical object to the four corners of which he assigned the four bases: adenine (A), tyrosine (T), guanine (G) and cytosine (C). The sequence GGCATCGTTGAA (the four initial codons of A chain of human insulin), which is depicted in Fig. 1, is obtained by starting at the center of the square and moving towards the corner of the square assigned as G (G being the first base in the DNA sequence of A chain of human insulin) and stopping half way toward the corner. From this point one continues in the direction of the corner having the label of the second base and again one stop at the half way. The process continues till all bases have been represented as spots within the interior of the square. As a result one obtains a 2-D “map” of dispersed spots, which may appear as being at random but are not) as graphical representation of DNA. In order to illustrate construction of such representation we have added arrows, which Jeffrey

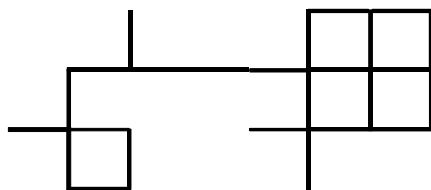
does not use. Let us mention two important properties of this particular graphical representation of DNA: (1) It is highly compact (allowing depicting DNA sequences having 100,000 bases); and (2) There is no loss of information associated with this construction, that is, given the final “map” one can always reconstruct the underlying DNA sequence.



**Figure 1.** The construction of the locations of initial 12 bases of human first exon of  $\beta$ -globin gene according to Jeffrey’s algorithm which uses the interior of a square for graphical representation of DNA.

In contrast the graphical representation of Nandy, illustrated in Fig. 2 on the full A-strand of human insulin,

even though it does not take too much space is in general neither compact nor it allows reconstruction, as is accompanied with some loss of information. However, it has the advantage of allowing visual inspection of such 2-D representations which facilitate recognizing, albeit qualitatively, the similarities and dissimilarities between different DNA sequences.



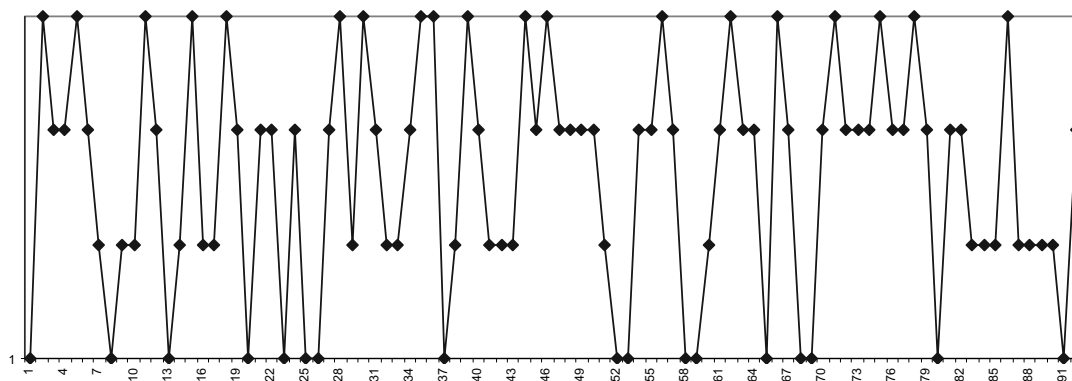
**Figure 2.** Graphical representation of A strand of human insulin according to Nandy's algorithm

Nandy used the coordinate grid of the Cartesian system assigning to the negative and the positive directions on the x-axis bases A and G respectively, and to the negative and the positive directions on the y-axis bases T and C, respectively. The 2-D representation of DNA is obtained as walk over the integer points on the coordinate grid, moving each time in the direction dictated by the considered base in the DNA sequence.

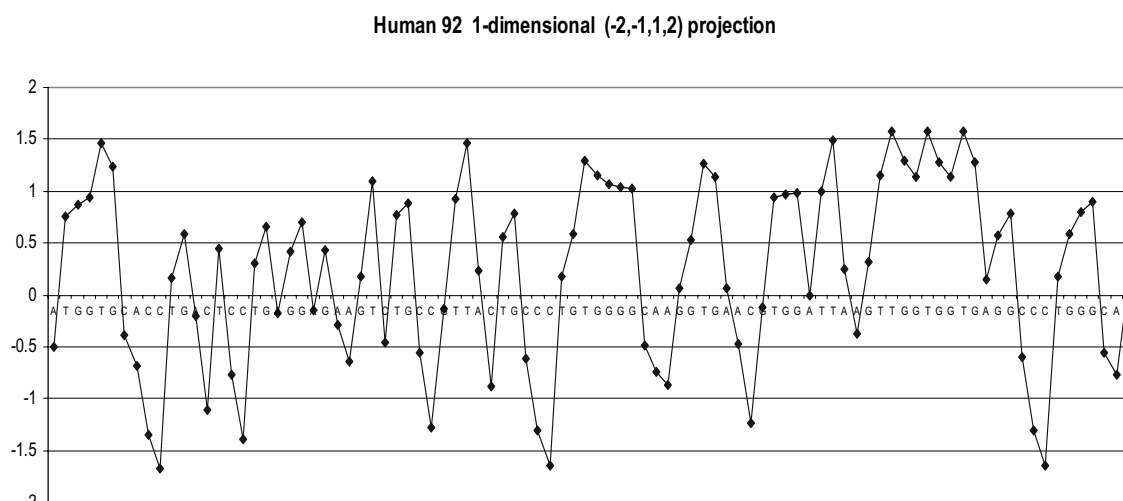
More recently<sup>10</sup> it has been realized that graphical representations of DNA can also be numerically characterized, which thus has upgraded graphical representations of DNA from a qualitative to a quantitative level. Moreover, not only that in this way hitherto qualitative representations of DNA could be characterized by quantitatively, but in addition it was possible to recover the information lost during the construction of the path of a walk in  $(x, y)$  plane. Apparently that this aspect of numerical characterization of DNA has not been immediately recognized is visible from efforts made to modify the approach of Nandy and reduce the accompanying loss of information.<sup>11–17</sup>

Introduction of matrices for analytical representation of graphical DNA images, which has led to their numerical characterization, indicated that for quantitative characterization of DNA are no longer so critical graphical representation, which only serves as an intermediate for construction of numerical representations of DNA. Dozens novel graphical and non-graphical representations of DNA followed, mainly from two research circles, our own,<sup>18–30</sup> and a school of Chinese scholars.<sup>31–51</sup> For a recent review on development and applications of all these graphical and non-graphical representation of DNA one should consult work of Nandy, Harle and Basak.<sup>52</sup> These novel constructions introduced a number of different underlying concepts, either by assuming different geometrical objects used as template, or using different properties of the nucleic acids for their classification, and resulted in diverse graphical representations of DNA. A smaller class of these novel graphical representations of DNA can be called “spectrum-like,” in view that their graphical representation has appearance of a spectrum given by a wavy curve, which extends horizontally but is confined vertically to a finite range. Spectrum-like curve can result from depicting a 1-dimensional data or 2-dimensional data. In the case of 1-D data the DNA sequential entries are given as a single numerical value, thus the “peaks” (y-coordinate) of the spectra are uniformly distributed along the x-coordinate. In the case of 2-D data the DNA sequential entries are represented by a pair of numbers  $(x, y)$ , thus “peaks” are no longer necessarily uniformly distributed along the x-coordinate axis. In other words 1-D “spectra” are defined solely by their amplitudes, while 2-D “spectra” are characterized by the “amplitude” and “frequency.”

In Fig. 3 we illustrate 1-D “spectrum-like” graphical representation of the first exon of human  $\beta$ -globin gene based on the approach outlined in ref. [14], in which to A (adenine), G (guanine), T (thymine) and C (cytosine) are assigned four horizontal lines on



**Figure 3.** Spectrum-like graphical representation of the first exon of human  $\beta$ -globin gene.



**Figure 4.** An alternative 1-D spectrum-like representation of the first exon of human  $\beta$ -globin gene outlined in ref. [22]

which, at equal intervals, are placed sequentially spots (points) corresponding to nucleotides as they appear in the DNA sequence. In Fig. 4 we show an alternative 1-D spectrum-like representation, outlined in ref. [22] of the same DNA obtained by assigning first to A, C, G, and T the values -1, -2, +1, and +2 respectively, starting from zero and using the algorithm of Jeffrey to depict as y-coordinate the location of DNA bases at a uniform step along the x-axis. There is some similarity between the 1-D representation of the same DNA shown in Fig. 3 and Fig. 4: They both use y-coordinate to indicate the “amplitudes” of the spectrum, but in Fig. 3 there are only four distinct y-values, while by the nature of Jeffrey’s algorithm no two spots in Fig. 4 have the same y-coordinate!

In this article we will introduce novel 1-D “spectrum-like” graphical representation of DNA, which will be based on codons, the triplets of nucleic acids in exon portions of DNA, rather than on individual nucleotides. There are two reasons for doing this: (1) Codons determine protein sequence, which thus determines the degree of similarity among proteins; and (2) Graphical representation based on codons are not only shorter (by a factor of three) but can be immediately extended for construction of 1-D protein “spectrum-like graphical representation, when the underlying RNA codons are known, as already outlined in ref. [18]. Moreover, even when codons determining a particular protein are not known (which is going to be often the case) one can arrive at 1-D spectrum-like representations of proteins using the Virtual Genetic Code,<sup>25</sup> which assigns to each amino acid a single triplet of nucleic bases. Similarity among DNA based on characterization of codons rather than individual bases may have some advantages. For example, consider two possibilities: in the first case a single base in DNA

sequence is mutated; and in the second two adjacent bases belong to the same codon are mutated. When one considers the similarity of so mutated DNAs with the original DNA sequence the second case would lead to a greater departure from the original sequence. However, the both cases may be accompanied by a change of a *single amino acid* in the protein generated, thus indicating the same degree of similarity.

## 2. Graphical Representation of Codons

It is not accidental that it took over 20 years after the first graphical representations of DNA to arrive at the first graphical representations of proteins.<sup>25, 26, 53-55</sup> Already the multiplicity of assigning the four nucleic bases between different direction along the coordinate axes, choosing one among several alternative assignments of the four bases to the four corners of a square, or four horizontal lines has pointed to dilemmas requires one to make some arbitrary decisions in selecting among seemingly equivalent geometrical choices when considering graphical representations of DNA. The number of such possibilities increases exponentially when instead of four bases one considers 20 amino acids, and becomes prohibitive when one wish to extends such considerations to 64 codons. Hence, one needs some *singular* “solution” that appears plausible, “logical” and “structural” if possible. Even a drastic reduction of available alternative routes to the goal – graphical representation of proteins and DNA based on 64 codons – would not be enough!

In the case of proteins one route was based on a scheme in which 20 amino acids have been ordered on circumference of the unit circle alphabetically (based on three letter codes).<sup>27</sup> Alphabetic order is essentially the

same as a “random order,” hence devoid of any inherent *structural* component. In the case of codons even such self-evident singular “solution” would not be of interest in view that alphabetic ordering of codons will not necessarily group codons coding the same amino acid in vicinity. It is true that when one consider comparisons among proteins and searches for protein alignment it need not be critical in which order are for example the 20 amino acids are placed on the circumference of a circle, as long as the same ordering of amino acids is maintained. This has been demonstrated by Randić, Butina and Zupan,<sup>27</sup> who considered alignment between two shorter proteins of yeast *Saccharomyces cerevisiae* ordering first amino acids alphabetically with respect to their three letter code and a single letter code. In both cases very similar plot of the difference between the amino acids in the same sequential position in protein I and protein II was obtained.

Before addressing the problem of alignment let us point out one important advantage of numerical characterization of protein sequences, whether based on graphical or non-graphical representations. In contrast to standard procedures for alignment of protein sequences based on computer packages and underlying algorithms (several of which will be mentioned in the next section), which only can apply to pair of proteins or a subset of proteins but cannot characterize a *single* protein sequence, numerical characterizations holds for single *individual* proteins. This is an important advantage of such representations, which allows proteins to be meaningfully catalogued and may facilitate protein classifications. Clearly numerical characterization of individual DNA sequence, or protein sequence will depend on ordering of bases or amino acids assumed when constructing the underlying geometrical objects used in the model. However, in subsequent comparative studies any of so adopted characterization will yield similar results which will mostly in the case of valid models be independent of the conventions used in ordering bases, amino acids, or codons.

### 3. Alignment

Alignment of protein sequences is one of the central topics of molecular biology. Protein sequence alignment is a standard tool for comparative study of proteins. It has been used to investigate evolutionary relationship among sequences that may point to a common ancestor. Several algorithms have been proposed since 1970 for protein alignment, starting with visual alignment known as dot plots, in which regions of similarity appear as diagonals within the dot plot matrix.<sup>56, 57</sup> In pair wise alignment one can search globally<sup>58, 59</sup> or locally<sup>60, 61</sup> for the best-matching taking into account mismatches in the alignment, which

correspond to mutations and which can be minimized by insertions and deletions. The common characteristic of all these computer-based alignment optimization scheme, including more recent ones, such as BLAST (Basic Local Alignment Search)<sup>62</sup> is adoption of scoring schemes, which assign positive score for match and negative score for mismatch and indels (short for insertion and deletions). While undoubtedly computer search will find optimal alignment, the solution will depend on assumed scoring, thus leaving the problem solved approximately.

Recently a fresh look at the old problem of protein alignment resulted in a novel rather elegant algorithm for protein sequence alignment which is devoid (at least for now) of scoring of matches, mismatches and gaps, which has been referred to as graphical alignment approach.<sup>63</sup> Admittedly, the novel approach is still under development and expansion.<sup>63-56</sup> Clearly, what we have here is a proof of new principle, rather than elaborate demonstration of indiscriminate useful application of the new scheme – which ten to twenty and more years from now, which is the age of most currently used computer based algorithms, may also evolve in competitive protein alignment procedure, or at least be in part incorporating in some of the existing protein alignment schemes. Notably the current approach has yet to address the issue of insertions and deletions, even though as it appears now there should be no problems with detection of insertion/deletion of a single base pair or expansion by only one base pair, and even shorter subsequences.

The basic novelty of the geometry-based graphical alignment approach consists in replacing the alphabetic (20 letter protein code) by numerical (20 number protein code). Once such a replacement is made one can apply simple meaningful arithmetic manipulations with sequences, such as subtracting one-variable numerical sequences of two proteins, that may also be shifted relative one to another in both directions by one or more steps.<sup>63</sup> Here one may not be necessarily interested in the magnitudes of the individual entries of the differences between two protein sequences, but in a binary output of such comparisons – whether two amino acids are the same or not. In this work we would like to report on similar comparisons of DNA sequences but rather than considering individual bases we would like to develop numerical characterization of DNA sequences based on 64 codons.

In order to reduce the number of “decision” steps that involve the large number of entries (64 codons), we will first place all 64 codons on an 8x8 Table of Codons.<sup>26, 28</sup> The 8x8 Table of Codons is a novelty introduced only recently obtained by applying the algorithm of Jeffrey,<sup>4</sup> but considering only triple of nucleotides and terminating the construction after the third nucleic acid.

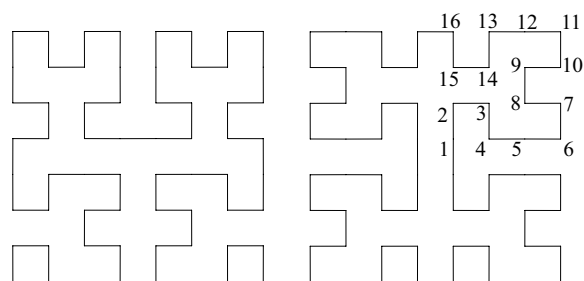
In this way one obtains unique location of all 64 codons (including the three “stop” triplets, one of which has been recently identified as a codon for the 21<sup>st</sup> natural amino acid).<sup>66,67</sup> The distribution of the codons over the 8x8 table depends not only on the assignment of A, C, G, and T to the four corners of a square, the interior of which is used for representing DNA, but also on the algorithm used in the construction. The algorithm of Jeffrey gives the priority to the third amino acid, rather than to the first and second and this is less desirable in view that the first two bases dominate the genetic code. More recently the algorithm of Jeffrey has been suitably modified<sup>28</sup> so that distribution of codons dominates the first and the second base. We have here adopted so modified 8x8 Table which has been reproduced as Table 1 from ref. [28].

**Table 1.** The 8x8 Table of Codons used for construction of polar coordinates for 64 codons

CCC	CCT	CTC	CTT	TCC	TCT	TTC	TTT
CTA	CCG	CTA	CTG	TCA	TCG	TTA	TTG
CAC	CAT	CGC	CGT	TAC	TAT	TGC	TGT
CAA	CAG	CGA	CGG	TAA	TAG	TGA	TGG
ACC	ACT	ATC	ATT	GCC	GCT	GTC	GTT
ACA	ACG	ATA	ATG	TCA	TCG	GTA	GTG
AAC	AAT	AGC	AGT	GAC	GAT	GGC	GGT
AAA	AAG	AGA	AGG	GAA	GAG	GGA	GGG

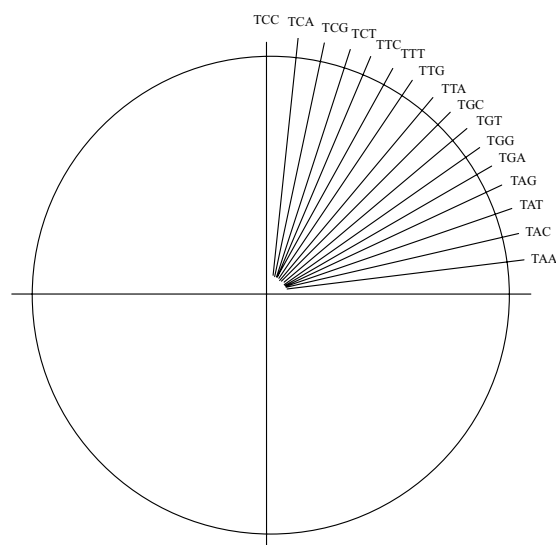
Here the 64 codons are placed so that codons having the same first nucleic acid are located in the same quadrant. As a consequence codons corresponding to the same amino acid, which often have same the first and the second base will be located in vicinity.

The next task is to “transfer” so ordered codons with as little as possible perturbation to the periphery of the unit circle, so that codons having the same first and the second base are also in the same neighborhood of the circle periphery. This is easy to state but not necessarily easy to accomplish. In order to arrive at one such solution, which will follow some rule rather than being arbitrary we will take advantage of the Hilbert fractal, that starts with 4 points and in the next stage involves 16, points and then 64 etc.<sup>68,69</sup> We will consider the third stage of the Hilbert fractal involving 64 vertices, which is shown in Fig. 5 at the left. At the



**Figure 5.** Hilbert fractal at the stage having 64 points (left) and its cyclic modification (right) with initial labeling of vertices.

right in Fig. 5 we show a modification of the Hilbert fractal which resulted in a closed cyclic fractal that not only passes through all 64 points located in the center of all cells of an 8x8 table, but grouped together codons having the same two initial bases. Thus the four codons 1-4 start with TA, the four codons 5-8 start with TG, the four codons 9-12 start with TT, and the four codons 13-16, which complete the first quadrant, start with TC. The 16 codons of the second quadrant start with C, the next 16 codons placed in the third quadrant start with A and the last group of 16 codons in the fourth quadrant start with G. We will now use the cyclic 64 points fractal to “transform” the entries of Table 1 into an ordered list of 64 codons on the periphery of the unit circle (that is, the circle of radius 1, placed in the origin of the Cartesian coordinate system).



**Figure 6.** The location of the first 16 codons of the first quadrant.

In Fig. 6 we have illustrated the process of the construction of unit circle of 64 codons on its periphery for the first 16 codons which belong to the first quadrant. The corresponding polar coordinates of these 16 codons are listed in Table 2.

**Table 2.** The polar coordinates for the first 16 codons arranged as outlined in Fig. 5 and Fig. 6 by placing codons on the periphery of the unit circle according closed walk of modified Hilbert fractal

TAA	0,098175
TAC	0,196350
TAT	0,294524
TAG	0,392699
TGA	0,490874
TGG	0,589049
TGT	0,687223
TGC	0,785398
TTA	0,883573
TTG	0,981748
TTT	1,079922
TTC	1,178097
TCT	1,276272
TCG	1,374447
TCA	1,472622
TCC	1,570796

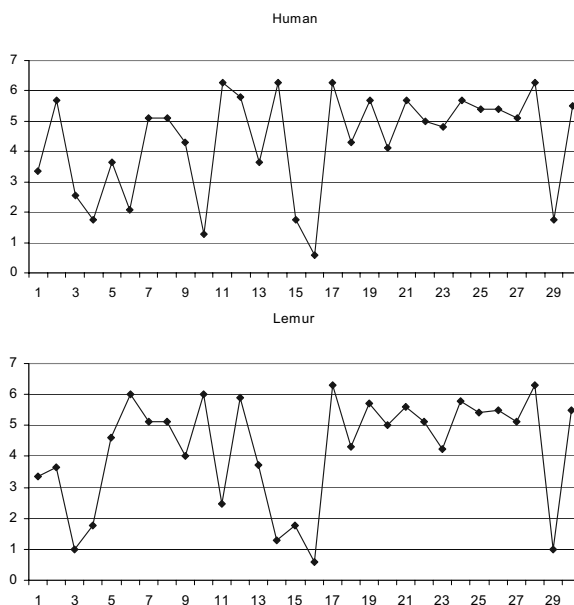
By arranging 64 codons on the periphery of the unit circle we have prepared the geometric basis for our graphical representation of DNA using triplets as building blocks, which is particularly suited for comparative study of exons.

#### 4. Illustration

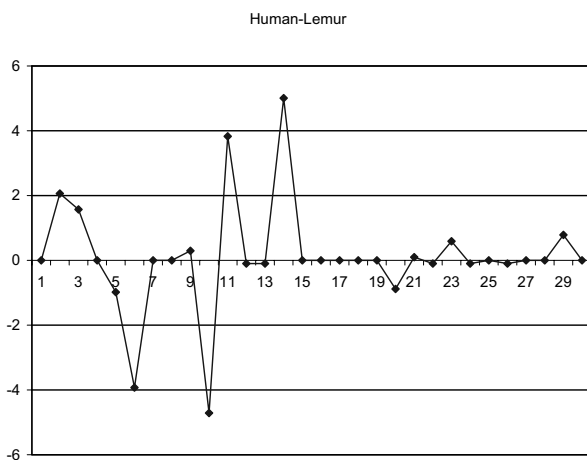
We will illustrate the present 1-D spectrum-like representation of exon portions of DNA based on codons on the triplets of nucleic acids occurring in the first exon of human and lemur  $\beta$ -globin gene, which are listed in Table 3 together with the corresponding polar coordinates for codons present. In Fig. 7 we have plotted the corresponding numerical sequences of Table 3, which result in spectrum-like “signatures” for the two individual DNA sequences considered. In the case of the first exon of human and lemur  $\beta$ -globin gene already visual inspection of the graphical representations shows that the two exons are fairly similar. A closer comparison of the two  $\beta$ -globin exons is facilitated by plotting the difference of the polar coordinates of the corresponding codons of two exons shown in Fig. 8. As one can from Fig. 8 there are 13 places at which the codons coincide when the difference in polar coordinates of corresponding codons equal zero and the result is plotted as spots on the x-axis. Observe that in addition we have six places where the difference between the polar coordinates of the corresponding codons of the two exons is quite small. This means that the corresponding pair of codon, one belonging to human and the other to lemur DNA are adjacent on the circumference of the unit circle. Because we have grouped the codons in the four quadrants according to

**Table 3.** The codons of the first exon of human and lemur  $\beta$ -globin gene and the associated polar coordinates (in radians). The last column shows the difference of coordinates of corresponding triplets.

Human	Lemur	Difference
ATG	ATG	0
GTG	ACT	2.0616710
CAC	TTG	1.5707960
CTG	CTG	0
ACT	AGT	-0.981750
CCT	GCT	-3.927490
GAG	GAG	0
GAG	GAG	0
AAG	AAT	0.294524
TCT	GCT	-4.712390
GCC	CAT	3.828816
GTT	GTC	-0.098180
ACT	ACC	-0.098170
GCC	TCT	5.006913
CTG	CTG	0
TGG	TGG	0
GGC	GGC	0
AAG	AAG	0
GTG	GTG	0
AAC	GAT	-0.883580
GTG	GTA	0.098175
GAT	GAG	-0.098180
GAA	AAA	0.589049
GTG	GTT	-0.098170
GGT	GGT	0
GGT	GGC	-0.098170
GAG	GAG	0
GCC	GCC	0
CTG	TTG	0.785398
GGC	GGC	0



**Figure 7.** Graphical representation of the first exon of  $\beta$ -globin gene of human and lemur respectively based on polar coordinates of Table 2.



**Figure 8.** Alignment of codons of the first exon of  $\beta$ -globin gene of human and lemur.

the first nucleotide, and in addition have placed codons having same the first and the second base in vicinity, the small differences in the difference plot belongs to triplets that will differ only in the third base. Hence, therefore some of the “peaks” in the difference plot of very magnitude may code for the same amino acid, thus even though triplets would be different the resulting protein will be the same. For example, from Table 3 we see that beside the 13 codons that are identical for both human and lemur there are in addition six codons for which the difference in the corresponding polar coordinates attain the smallest possible value:  $\pm 0.09818$ . We have listed all the six cases of the smallest possible non-zero difference below, five of which as one can see code for the same amino acid:

Human	Lemur	Corresponding amino acids
GTT	GTC	valine-valine
ACT	ACC	threonine-threonine
GTG	GTA	valine-valine
GAT	GAG	asparagine-glutamic acid
GTG	GTT	valine-valine
GGT	GGC	glycine-glycine

Hence, the two exons will produce 18 identical amino acids (close to 2/3 of the total of 30 amino acids associated with the first exon). We may conclude therefore that the visual representation of the differences between the graphical “signatures” of human and lemur first exon of  $\beta$ -globin gene (illustrated in Fig. 8) represents suitable methodology for examination of the degree of similarity between DNA sequences considered.

## 5. Conclusions

We have outlined a simple numerical representation of DNA based on codons, in which to each triplet of

nucleic acid a single numerical parameter is assigned. The numerical values of codons represent the polar coordinate when 64 codons are distributed uniformly on the circumference of the unit circle. Codons have been arranged on the circle periphery in such a manner that neighboring locations on the periphery of the unit circle belongs to codons having the first and often also the second nucleotide in common. As a consequence DNA sequences having only few mutations (substitutions) will be represented by numerical sequences that will not be very different, particularly when mutation involves the third nucleic base. Graphical illustration of such DNA sequences lead to a “spectrum-like” representations, which may allow visual recognition of similarities in different sequences. The comparison can be further facilitated by plotting the numerical difference of two sequences, which will show domains of the zero difference (portions of the sequence which are in full alignment) and also domains of “almost zero difference.” The latter correspond to triplets in which the difference is due to the third base, which because of the degeneracy of the Genetic Code, may still code for the same protein. An advantage of the proposed graphical approach is that it associates with triplets constant single variable coordinate at no loss of information on the DNA sequence considered, which can always be fully recovered from its numerical sequence.

## 6. Acknowledgments

The author wishes to express thanks to Professor Jure Zupan and Dr. Marjana Novič of the Laboratory for Chemometrics of the National Institute of Chemistry, Ljubljana, Slovenia for the warm hospitality and to the Ministry of Science of Slovenia for the financial support within the project P1-017: “Modeling of relationship between chemical structure and properties: QSAR–QSPR.”

## 7. References

1. E. Hamori, *Nature* **1985**, *314*, 585–586.
2. M. A. Gates, *J. Theor. Biol.* **1986**, *119*, 319–328.
3. H. J. Jeffrey, *Nucleic Acid Res.* **1990**, *18*, 2163.
4. C. A. Pickover, *J. Mol. Graphics* **1992**, *10*, 2.
5. A. Nandy, *Curr. Sci.* **1994**, *66*, 309–314.
6. P. M. Leong and S. Mogenthaler, *Comput. Appl. Biosci.* **1995**, *12*, 503–507.
7. J. Ninio and E. Mizraji, in: Pickover, C. A. ed., *Visualizing Biological information*. World Scientific: Singapore, **1995**, pp. 33–42.
8. Y. K. Huen, in: Pickover, C. A. ed., *Visualizing Biological information*. World Scientific: Singapore, **1995**, pp. 165–182.
9. B. E. Blatsdell, A. M. Campbell and S. Karlin, *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93*, 5854–5859.

10. M. Randić, M. Vračko, A. Nandy and S. C. Basak, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 1235–1244.
11. X. Guo, M. Randić, and S. C. Basak, *Chem. Phys. Lett.* **2002**, *350*, 106–112.
12. Y. Liu, X. Gou, J. Xu, L. Pan, and S. Wang, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 529–533.
13. X. Guo and A. Nandy, *Chem. Phys. Lett.* **2003**, *369*, 361–366.
14. S. S. T. Yau, J. Wang, A. Niknejad, C. Lu, N. Jin, Y. Ho, *Nucleic Acid Res.* **2003**, *31*, 3078–3080.
15. B. Liao, *Chem. Phys. Lett.* **2005**, *401*, 196–199.
16. B. Liao, M. Tan, K. Ding, *Chem. Phys. Lett.* **2005**, *414*, 296–300.
17. Y. Wu, A. W. Liew, H. Yan, and M. Yang, *M. Chem. Phys. Lett.* **2003**, *367*, 170.
18. M. Randić, M. Vračko, N. Lerš, and D. Plavšić, *Chem. Phys. Lett.* **2003**, *368*, 1–6.
19. M. Randić, M. Vračko, N. Lerš, and D. Plavšić, *Chem. Phys. Lett.* **2003**, *371*, 202–207.
20. M. Randić, M. Vračko, J. Zupan, and M. Novič, *Chem. Phys. Lett.* **2003**, *373*, 558–562.
21. M. Randić, *Chem. Phys. Lett.* **2004**, *386*, 468–471.
22. M. Randić and J. Zupan, *SAR & QSAR in Environ. Res.* **2004**, *15*, 191–205.
23. M. Randić, N. Lerš, D. Plavšić, S. C. Basak, and A. T. Balaban, *Chem. Phys. Lett.* **2005**, *407*, 205–208.
24. M. Randić, *Period. Biolog.*, **2005**, *107*, 415–422.
25. M. Randić, D. Vikić-Topić, A. Graovac, N. Lerš and D. Plavšić, *Period. Biolog.*, **2005**, *107*, 437–444.
26. Zupan, J.; Randić, M., *Advances in Quantum Chemistry* (special issue on Chemical Graph Theory, D. J. Klein, guest editor), in press.
27. J. Zupan, and M. Randić, *J. Chem. Inf. Model.* **2005**, *45*, 309–313.
28. M. Randić, *SAR & QSAR in Environ. Res.* **2004**, *15*, 147–157.
29. M. Randić, J. Zupan, and A. T. Balaban, *Chem. Phys. Lett.* **2004**, *397*, 247.
30. Randić, M.; D. Butina and J. Zupan, *Chem. Phys. Lett.* **2006**, *419*, 528.
31. M. Randić, A. T. Balaban, M. Novič, A. Založnik and T. Pisanski, *Period. Biolog.* **2005**, *107*, 403–414.
32. S. S-T. Yau, J. Wang, A. Niknejad, C. Lu, N. Jin and Y.-K. Ho, *Nucleic Acid Res.* **2003**, *31*, 3078.
33. P. He, J. Wang, *Internet Electron. J. Mol. Des.* **2002**, *1*, 668–674.
34. Li, C.; Wang, J., *Comb. Chem. High T. Scr.*, **2004**, *7*, 23–27.
35. Wu, Y.; Liew, A. W.; Yan, H.; Yang, M., *Chem. Phys. Lett.* **2003**, *367*, 170–176.
36. J. Song, H. Tang, *J. Biochem. Biophys. Methods* **2005**, *63*, 228–239.
37. Y. Yao, B. Liao, T. Wang, *J. Mol. Struct. (Theochem)* **2005**, *755*, 131–136.
38. C. Li, J. Wang, *Combinatorial Chem. & High Throughput Screening* **2003**, *6*, 795–799.
39. B. Liao, T. Wang, *J. Comput. Chem.* **2004**, *25*, 1364–1368.
40. B. Liao and K. Ding, *J. Comput. Chem.* **2005**, *26*, 1519–1523.
41. J. Wang and Y. Zhang, *Chem. Phys. Lett.* **2006**, *423*, 50–53.
42. Y. Yao, T. Wang, *Chem. Phys. Lett.* **2004**, *398*, 318–323.
43. C. Li, J. Wang, *Combinatorial Chem. & High Throughput Screening* **2004**, *7*, 23–27.
44. Y. Yao, X. Nun, T. Wang, *Chem. Phys. Lett.* **2005**, *411*, 248–255.
45. C. Yuan, B. Liao, T. Wang, *Chem. Phys. Lett.* **2003**, *379*, 412–417.
46. B. Liao, T. Wang, *J. Mol. Struct. (Theochem)* **2004**, *681*, 209–212.
47. B. Liao, Y. Zhang, K. Ding, T. Wang, *J. Mol. Struct.* **2005**, *717*, 199–203.
48. W. Zhu, B. Liao, K. Ding, *J. Mol. Struct.* **2005**, *717*, 193–198.
49. F. Bai, W. Zhu, T. Wang, *Chem. Phys. Lett.* **2005**, *408*, 258–263.
50. R. Chi, K. Ding, *Chem. Phys. Lett.* **2005**, *407*, 63–67.
51. B. Liao, T. Wang, *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1666–1670.
52. A. Nandy, M. Harle, S. C. Basak, ARKIVOC (electronic journal) (in press).
53. Randić, M.; M. Novič, Vikić-Topić, D., D. Plavšić, *SAR & QSAR in Environ. Res.* (in press, November 2006).
54. Randić, M.; Zupan, J.; Vikić-Topić, D., *J. Mol. Graphics & Model.* (press).
55. M. Randić, *Chem. Phys. Lett.* (submitted).
56. Gibbs A. J. and McIntyre, G. A. *Eur. J. Biochem.* **1970**, *16*, 1–11.
57. Staden, R., *Nucl. Acid Res.* **1982**, *10*, 2951–2961.
58. Smith, T. F.; Waterman, M. S., *J. Molecular Biology*, **1981**, *147*, 195–197.
59. <http://bioweb.pasteur.fr/interfaces/water.html>
60. Needleman, S.; Wunsch, C. D., *J. Molecular Biology*, **1970**, *48*, 443–453.
61. <http://bioweb.pasteur.fr/seqanal/interfaces/needle.html>
62. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. And Lipman, D. J., *J. Mol. Biol.* **1990**, *215*, 403–410.
63. M. Randić, *J. Math. Chem.* (submitted).
64. M. Randić and J. Zupan, Vikić-Topić, D. Plavšić, *Chem. Phys. Lett.* **2006**, *431*, 375–379.
65. M. Randić, *Chem. Phys. Lett.* (submitted).
66. I. Chambers, J. Frampton, P. Goldfarb, N. Affara, W. McBain, and P. Harrison, *EMBO J*, **1986**, *5*, 1221–1227.
67. F. Zinoni, A. Birkmann, T. Stadtman, A. Boeck, *Proc. Natl. Acad. Sci. U. S. A.* **1986**, *83*, 4650–4654.
68. D. Hilbert, *Math. Ann.* **1891**, *38*, 459–460.
69. V. Batagelj (Dept. of Mathematics, Univ. of Ljubljana) has developed software that draws Hilbert fractals at ever increasing stage.



## Povzetek

Preučevali smo novo 1-dimenzionalno reprezentacijo DNA, ki temelji na grafičnih predstavitev 64 tripletov jedrskih kislin na krožnici s polmerom ena. Z uporabo polarnih koordinat 64 kodonov (izraženih v radianih) pretvorimo štiričrkovno sekvenco DNA v številčno sekvenco z ne več kot 64 različnimi zapisi. Spektro-podobno reprezentacijo DNA dobimo z upodobitvijo 1-dimenzionalne reprezentacije (nanešene na ordinato), pri čemer abscisa predstavlja tekoči zaporedni indeks. Nova reprezentacija DNA ima nekaj prednosti pred ostalimi spektro-podobnimi eno- in dvo-dimenzionalnimi reprezentacijami. Uporablja namreč iste koordinate za iste kodone, s čemer se izognemo vsakokratnemu novemu izračunu koordinat, kar je bilo značilno za Jeffrey-jev algoritem in izpeljanke grafičnih reprezentacij DNA iz tega algoritma.