

*Review*

## CHEMICAL MODELING USING GRAPHS: FROM ADJACENCY MATRIX TO RANDIĆ CODING OF DNA

**Alexandru V. Botez, Mihai V. Putz** \*

Laboratory of Computational and Structural Physical-Chemistry for Nanosciences and QSAR, Department of Biology – Chemistry, Faculty of Chemistry, Biology, Geography, West University of Timisoara, 16 Street Pestalozzi, Timisoara 300115, Romania

### ABSTRACT

Graph Theory offers tools for a wide area of fields: mathematics, physics, chemistry, biology, biochemistry, bioinformatics and computer science. Graphical representations have been used not only for comparative studies of genetic material, but also to visualize real or fictional relationships and interactions between vertices. This review aims to highlight some of the most frequently used transdisciplinary applications of graph theory and to present the steps necessary to build such graphs. The development of graphical representations enabled the comparisons of similarities and differences between long DNA sequences.

**Keywords:** graph theory, graphical modeling, transdisciplinary applications, virtual genetic code, brain connectivity

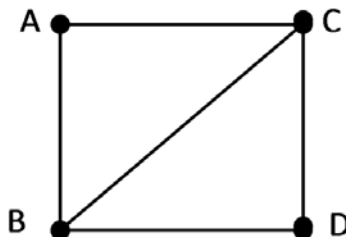
### 1. INTRODUCTION

The history of Graph theory begins in 1735 with Leonhard Euler and the Seven Bridges of Königsberg puzzle, in which one must cross each bridge only once to arrive at the starting point. Euler demonstrated that the puzzle has no solution [1].

A graph (Figure 1) is an ordered pair  $G = (V, E)$  comprising of  $V$  vertices or nodes together with  $E$  edges or lines. A simple graph is undirected [2].

---

\* Correspondent author: Tel: 004(0)-256-592-638, Fax: 004(0)-256-592-620, E-mail: mv\_putz@yahoo.com

**Figure 1:** A basic graph

Graph Theory [2] started 35 years ago as a branch of mathematics that studies geometry [3], function-spaces theory [4], geomorphology [5] and combinatorial matrix theory [6]. Graph Theory facilitated the development of Chemical Graph Theory, which enriched QSAR studies and drug modeling [7].

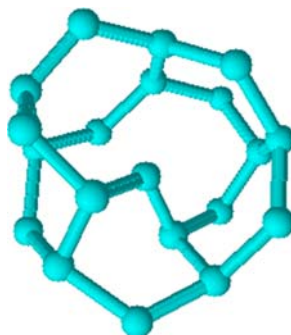
Although Hamori was the one who initiated graphical representations of DNA over two decades ago [8], the first graphical representation of a protein emerged only in the last decade [9].

## 2. CHEMICAL GRAPH MODELING BASIS

### 2.1. Types of Chemical Graphs

Types of Graphs are widely presented in literature in the following forms [10, 11]:

- Undirected graph - the edges of an undirected graph have no orientation;
- Directed graph - or digraph is an ordered pair  $G = (V, E)$  with  $V$  as a set of vertices or nodes and  $E$  as a set of ordered pairs of edges or arrows;
- Mixed graph - a graph  $G$  which have both directed and undirected edges;
- Multigraph - a graph containing multiple edges and/or loops;
- Quiver - or "multiDigraph" is a directed multigraph which can have more than one arrow from a given source to a given target;
- Simple graph - an undirected graph that has no loops (edges connected at both ends to the same vertex) and no more than one edge between any two different vertices;
- Weighted graph - a graph in which a number (weight) is assigned to each edge. Such weights might represent: costs, lengths or capacities, depending on the specific issue;
- Regular graph - a graph where each vertex has the same number of neighbors;
- Complete graph - connects each vertex to each edge;
- Finite and infinite graphs - a graph  $G = (V, E)$  such that  $V$  and  $E$  are finite sets. An infinite graph is one with an infinite set of vertices or edges or both;
- Subgraph - a subgraph  $G_0$  of  $G$  is a graph whose vertices and edges are contained in the graph  $G$ ;
- Möbius graph - a special kind of simple graph predicted by Heilbronner [12] in the form of a ladder or belt, with the ends connected in a half twist. An important property of the Möbius belt is that it has only one side in 3D space [13] (Figure 2). It was first synthesized in 2003 [14].

**Figure 2:** Möbius type cyclacene zigzag belt

## 2.2. Types of Graph Matrices

Among the most popular matrices used, we mention [15]:

- Adjacency matrix:  $A=A(G)$ , with  $N$  vertices, yields a  $N \times N$  symmetric matrix]:

$$[A]_{ij} = \begin{cases} 1 & \text{if } i \neq j \text{ and } e_{ij} \in E(G) \\ 0 & \text{if } i = j \text{ or } e_{ij} \notin E(G) \end{cases} \quad (1)$$

- Adjacency count matrix:  $DEG=DEG(G)$ , is the number of edges attached to each vertex:

$$[DEG]_{ij} = \begin{cases} \deg(v_i) & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \quad (2)$$

- Distance matrix:  $D=D(G)$ , which can be used to calculate the Wiener index ( $W$ ) and each vertices' contribution to the Wiener index ( $W$ ):

$$[D]_{ij} = \begin{cases} d_{ij} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \quad (3)$$

- Laplacian matrix:

$$[L]_{ij} = [DEG]_{ij} - [A]_{ij} \quad (4)$$

- X matrix:

$$[X]_{ij} = \begin{cases} (\deg_{ij})^{1/2} & \text{if } e_{ij} \in E(G) \\ 0 & \text{if } e_{ij} \notin E(G) \end{cases} \quad (5)$$

- Reciprocal distance matrix [15]:

$$R=R(G) \\ [RD]_{ij} = [1/D]_{ij} \quad (6)$$

- Detour matrix: the longest path between 2 vertices:

$$[\Delta]_{ij} = \begin{cases} \max(p_{ij}) & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \quad (7)$$

- Wiener matrices:  $W_e$ ,  $W_p$  and  $W_\Delta = W_p - W_e$  for acyclic graph:

- Wiener-edge matrix: the product of fragment vertices resulted from the cutting one edge;

$$[W_e]_{ij} = N_{i,e} \times N_{j,e} \quad (8)$$

- Wiener-path matrix: obtained by populating the zero values of the  $W_e$  matrix with the number of edges of either side of a considered path ( $V_{i,p}V_{j,p}$ ), including the path ( $V_{i,p}V_{j,p}$ );
- Wiener matrix:  $W_\Delta = W_p - W_e$ , giving the edge-graph contribution to the Wiener index (different from the distance contribution to Wiener);
- Combinatorial matrices:  $D_p, D_\Delta$ :
  - $D_p$ : the distance-path matrix is based on the distance matrix, but counts all possible paths formed inside of the shortest path between two vertices and allows direct calculation of the hyper-Wiener index;
  - $D_\Delta = D_p - D$ : gives the non-Wiener values of the hyper-Wiener index;
- Hosoya path matrix: a  $N \times N$  matrix for acyclic graphs built by cutting an edge  $ij$  of the graph and calculating the Hosoya index of the resulting subgraphs [16];
- Hosoya index: also known as the  $Z$  index, is the total number of groups ( $k$ =order of the group, if  $k=0$ ,  $Z$  is always 1) of non adjacent edges; was proposed to predict the order of boiling points for alkane isomers [17]. The sum of this matrix represents the path-Hosoya-Wiener index. Other such matrices are Hosoya edge matrix and Hosoya vertex matrix:

$$Z = \sum_{k=0}^{\lfloor V/2 \rfloor} p(G; k) \quad (9)$$

### 3. DISCUSSIONS

Graphs are an intrinsic part of understanding connectivity of networks and pathways, be it social, migration, information, biochemical or neurological. Here we survey some of them.

#### 3.1. Computer Science

Applications of graphs in computer science have led to the development of algorithms for solving problems [11, 18], such as:

- Shortest path algorithm in a network;
- Finding a minimum spanning tree;
- Finding graph planarity;
- Algorithms to find adjacency matrices;
- Algorithms to find the connectivity matrices;
- Algorithms to find the cycles in a graph;
- Algorithms for building search engines;
- Ordering of timed events and timed table scheduling;
- Algorithms for computer network security;
- Algorithms for fingerprint technology;
- Algorithms for decision making;

The most popular application of graph theory in computer science is the creation of the World Wide Web, which is nothing else than a large directed graph [19].

### 3.2. Physics

The crystal structure of compounds can be represented by graphs for the configuration of the face centered lattice [20]. Graphs can also be used for the characterization of molecular dynamics by clustering and geometrical rearrangements during crystallization [21]. Among the newest trends is NMR crystallography of zeolites in which the graph is built atom-by-atom [22]. This method is capable of characterizing crystals even with a low degree of periodicity, unlike diffraction techniques [23].

### 3.3. Chemistry

The most popular application of graph theory in chemistry is the Hückel molecular orbital theory [24].

Graph theory has been widely used in chemistry to make predictions on the compounds behavior of which we mention some:

- predicting diamagnetic properties of inorganic compounds for QSAR studies [25];
- predicting boiling points of alkanes [26] and their isomers [17];
- predicting branching of alkanes [27] and their thermodynamic stability [28];
- canonical naming of large clusters of isomers [29, 30];
- chemical reaction networks of simple and metabolic pathways [31];
- representation of three-dimensional molecular structures as basic graphs [32];
- predicting molecular similarity [33, 34];
- predicting topological energies and chemical stability of carbon structures [35, 36];

### 3.4. Biology

On the macro-scale, graph theory is key for the construction of migration routes of populations, models for human urbanization [37] and genealogical trees [38].

On the nano-scale, graph theory enhances QSAR studies and drug design by predicting drug availability interactions [39], toxicological [26] and carcinogenetic properties [40, 41].

### 3.5. Brain Topography

Graph theoretical modeling of brain connectivity has enhanced the knowledge of the brain's structure and functional systems as its complex networks are characterized by small-world topology [42], modularity [43] and highly connected hubs [44].

Human brain connectivity has been studied using functional MRI [45], diffusion tensor imaging tractography [46], magneto- and electroencephalography [47].

From the construction of the neural graph, the following observations have been made [43, 48-50]:

- the extent of a node (vertex) depends on the number of links to the rest of the network;
- the distance between nodes (edges) that has to be traversed is called a path;
- high node connectivity and short paths translates to high efficiency of parallel information transfer and fast processing;
- long paths offer fast transfer of information between distant dense nodes;
- dense populations of nodes follow hierarchical clustering and they are called hubs or clusters;
- hubs take on different roles in information processing and they are classified in modules;

Analyses of structural networks offer valuable information about the brain architecture, nevertheless, neurophysiological dynamics are hard to elucidate for a healthy individual [51]. When a normal brain is used as a control sample and compared to those with many neurological and psychiatric disorders such as autism, ADHD [52] and schizophrenia [53], it has been studied and observed that they are caused by node dysconnectivity [54].

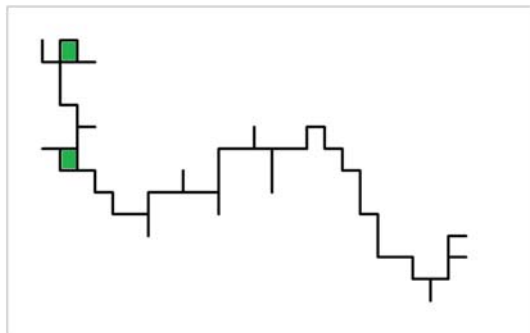
Great differences have been observed in the brain networks of the sick, still, a certain degree of dissimilarity can also be observed between healthy individuals. Additionally, complex brain networks are shared between individuals of different species [55].

### 3.6. Proteomics: The Randić Virtual Genetic Coding

The Virtual Genetic Code is a hypothetical string of nucleic acid bases (A, C, U, G) used to compile an existing protein, by assigning each amino acid (aa) an unique codon. It is possible to reconstruct this hypothetical DNA, while the actual genetic sequence that produces the protein remains unknown [56]. This permits the comparison of similarities and differences of sequences though 2D mathematical objects. These 2D graphical representations do not need to be unique or even to permit the reconstruction of the initial DNA sequence.

Such 2D graphical representations of DNA are the random-walk plot that have the advantage of visualizing certain periodic patterns, but they also hides any repeating patterns [57, 58], as seen in Figure 3.

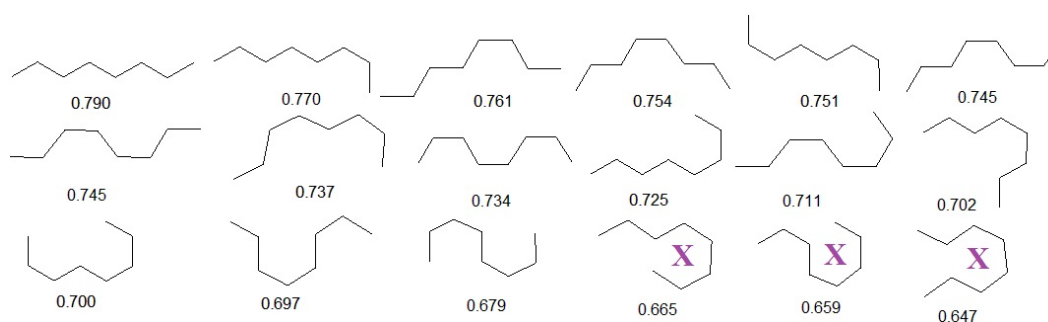
**Figure 3:** Graphical representation of the hypothetical RNA of the first 97 amino acids of the human ND6 protein, adapted from Randić [59]; see text for details



Numerical characterizations can accompany graphical representations of DNA sequences to allow both qualitative and quantitative representations of DNA [60].

This approach is used for characterization of bending and folding of molecular chain, through the use of eigenvalues from the distance/distance ( $D/D$ ) matrix, where  $(i, j)$  are two Euclidean distances between atoms, respectively through space and which are measured along the chain. Such eigenvalues of increasingly bent structures can be seen in the following graphical representation, as the Euclidean distances get smaller and the theoretical distances remain the same (Figure 4) [61].

**Figure 4:** Illustration of planar rotational isomers of eight-carbon polyene chains. While being discriminated upon the eigen-values of the extremes-connecting Euclidian distances, after [62]



The  $D/D$  matrix for proteins takes the shape of a compressed constant size  $20 \times 20$  matrix, which measures the sum of all distances for the selected amino acid, as seen in Table 1.

Beside the  $D/D$  matrix or the adjacency, in a similar manner, the construction of the adjacency-count  $(i, j)$  matrix is possible [64]. One proceeds by building a matrix of appropriate size and counting how many times  $i$  is the neighbor of  $j$ . These matrices are especially useful for the comparison of genetic material [62].

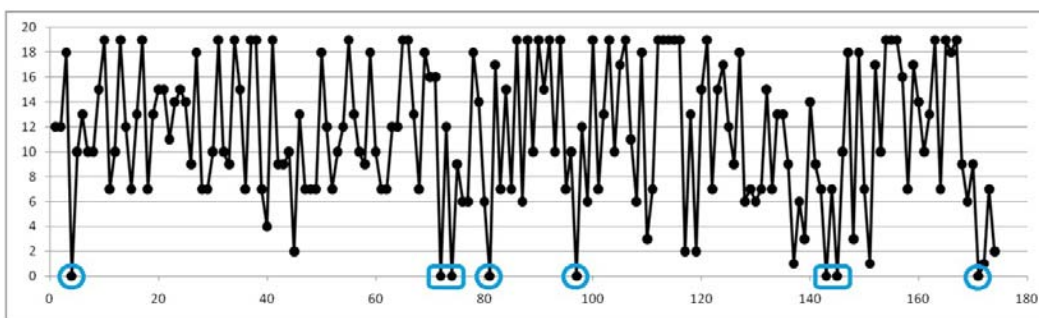
The comparisons of similarities and differences between long DNA sequences were otherwise almost impossible without the development of graphical representations. The basic methods in comparative studies of DNA use the three letter codons as its building blocks. While the correspondence between codons and amino acids is not unique, the differences in the coding of an amino acid can be assigned a squared weight value [62].

The metric comparison of DNA is achieved by assigning a number to each amino acid (Figure 5) and by aligning the 2 sequences. This approach coupled with codon weighing gives satisfactory results by itself, but it can be improved by shifting the sequence with one or more codons to give a superior alignment of DNA from different species [62].

For simplification of numeric sequence comparison, all similarities are assigned with 0-value while all differences with 1-value; this way, one can easily conclude the degree of similarity [62]. Similar to the alignment presented earlier, the spectrum alignment of DNA is both simple and effective. By assigning to each of the four bases, the numbers 1 to 4, one obtains an array of digits that can be processed in the same way already mentioned [65, 66], thus obtaining a DNA number sequence (Figure 6).

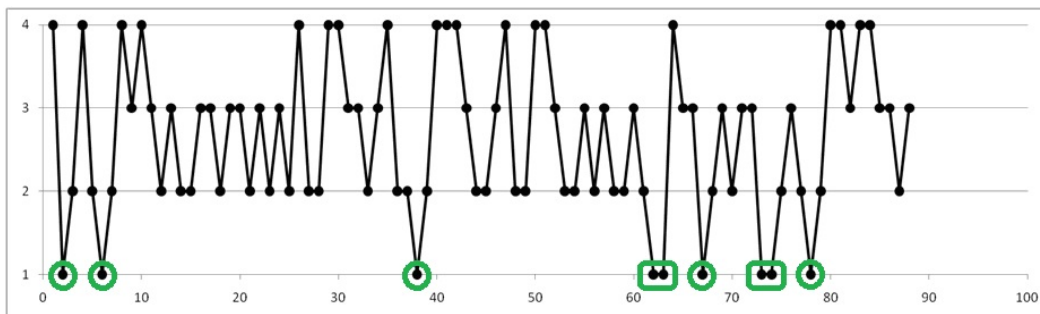
**Table 1:** Alphabetical order of the distance matrix for ND6 human protein, after [63]

	Ala	Arg	Asn	Asp	Cys	Gin	Glu	Glv	His	lie	Leu	Lvs	Met	Phe	Pro	Ser	Thr	Tip	Tyr	Val
A	0																			
R	113	0																		
N	35	78	0																	
D	70	45	36	0																
C	113	227	149	184	0															
Q	-	-	-	-	-	0														
E	72	47	38	6	184	-	-													
G	33	147	68	103	80	-	104	0												
H	-	-	-	-	-	-	-	-	0											
I	5	119	40	75	108	-	77	28	-	-										
L	68	182	103	138	46	-	CO	34	-	62	0									
K	50	162	84	118	67	-	169	17	-	44	22	0								
M	23	213	134	169	15	-	169	65	-	94	31	52	0							
F	44	172	93	128	55	-	128	24	-	53	9	13	41	0						
P	50	154	75	110	73	-	111	8	-	34	28	16	59	18	0					
S	68	182	103	138	46	-	138	34	-	63	2	21	31	10	29	0				
T	23	133	56	89	95	-	89	16	-	18	49	28	80	39	25	49	0			
W	80	40	47	12	193	-	8	112	-	85	147	126	178	137	120	147	97	0		
Y	43	157	78	113	71	-	113	10	-	38	25	8	56	15	8	25	24	122	0	
V	12	126	48	83	100	-	85	21	-	7	55	38	86	45	27	55	15	93	31	0

**Figure 5:** Spectrum representation of the *amino acids sequence* for the human ND6 protein, adapted from Randić [59]; see text for details

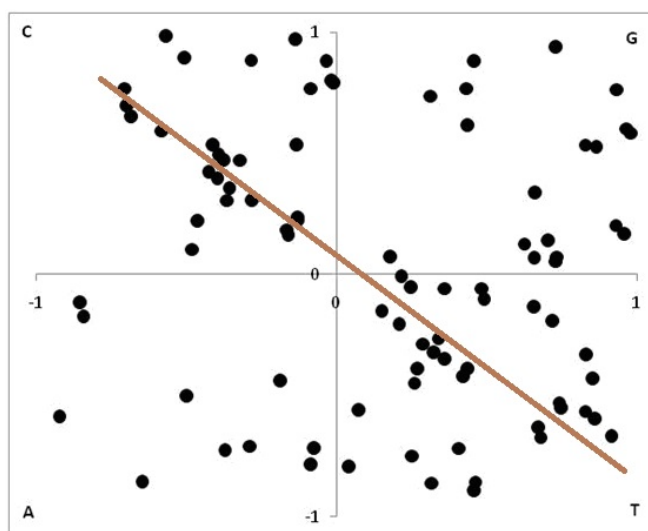


**Figure 6:** Spectral representation of the *hypothetical RNA* of the first 30 amino acids for the human ND6 protein, adapted from Randić [59]; see text for details



While DNA can be written as a long array of 4 letter (or numbers), after the transcription to RNA, this numbering can be maintained with a loss of information on its secondary structure [67], or the numbering can be increased to 8 or 12. The first set of 4 belongs to the unbound bases, the second belongs to the hydrogen bound bases, while the third belongs to the bases that are hydrogen bound to the second group. This smilingly large pool of numbers for only 4 letters faithfully translates the secondary structure of RNA from START to FINISH [62]. These number sequences allow writing and reading as well as a full reconstruction of the parent genetic material.

**Figure 7:** Graphical representation of Jeffrey's Magic Square for the hypothetical RNA of the first 30 amino acids for the human ND6 protein of Figure 6, adapted from Randić [59]; see text for details



Jeffrey's Magic Square is a highly compact 2D map created by assigning the four bases A, C, G and T to each corner of a square [68]. The sequence starts in the center of the square and moves half the distance to the designated corner of the first letter, then half the distance to

the next sequenced base, leading to a collection of points that represents that unique strand of genetic material (Figure 7). This method of DNA comparison which looks similar to Michael Barnsley's "Chaos Game" [69], is useful only for comparing small DNA sequences.

For the application of Jeffrey's Magic Square to proteins the shape of the representation changes from a square to an Icosagon. A circle with 20 points is the optimized representation of Jeffrey's modified square for proteins, even though it creates a more chaotic pattern [62].

Before the graphical representations of DNA and proteins, their mathematical processing offers all the necessary information for their characterization and comparison.

When graphical representations of DNA and proteins become too complex because of 3+ dimensional vectors, informational reprocessing has to take a different form. Beside the distance matrices presented earlier, assigning a vector to each of the 20 natural amino acids in the 20D distance matrix it will give the possibility of 20 equivalent directions (Table 2) to be exploited in the forms of walks in space (Table 3) and to construct the protein profile's (Figure 8) [70]. The same method can also be applied for nucleotides.

**Table 2:** Vector coordinates for the 20 natural amino acids, after [70]

1	<b>Ala</b>	<b>A</b>	000000000000000000 1
2	<b>Arg</b>	<b>R</b>	000000000000000000 10
3	<b>Asn</b>	<b>N</b>	000000000000000000 100
4	<b>Asp</b>	<b>D</b>	000000000000000000 1000
5	<b>Cys</b>	<b>C</b>	000000000000000000 10000
6	<b>Gin</b>	<b>Q</b>	000000000000000000 100000
7	<b>Glu</b>	<b>E</b>	000000000000000000 1000000
8	<b>Gly</b>	<b>G</b>	000000000000000000 10000000
9	<b>His</b>	<b>H</b>	000000000000000000 100000000
10	<b>Ile</b>	<b>I</b>	000000000000000000 1000000000
11	<b>Leu</b>	<b>L</b>	000000000000000000 10000000000
12	<b>Lys</b>	<b>K</b>	000000000000000000 100000000000
13	<b>Met</b>	<b>M</b>	000000000000000000 1000000000000
14	<b>Phe</b>	<b>F</b>	000000000000000000 10000000000000
15	<b>Pro</b>	<b>P</b>	000000000000000000 100000000000000
16	<b>Ser</b>	<b>S</b>	000000000000000000 1000000000000000
17	<b>Thr</b>	<b>T</b>	000000000000000000 10000000000000000
18	<b>Trp</b>	<b>W</b>	000000000000000000 100000000000000000
19	<b>Tyr</b>	<b>Y</b>	000000000000000000 1000000000000000000
20	<b>Val</b>	<b>V</b>	10000000000000000000

These vectors form the 20D path which is populated by the amino acids in the protein sequence. The path or walk, as the author calls it [70], is a cumulating process of all steps taken along the protein.

For the sequence: MMYALFLLSVGLVMGFVGFSL we calculate the end path: 31002033040030000001, but to obtain a clear array, all the points along its path have to be plotted. From the arrays' path other eigenvalues can be extracted, such as: the frequency ( $f_x$ ), the increment number ( $i=f_x + f_{x-1}$ ), the sum ( $s$ ) of "T" and the distance ( $s^{1/2}$ ) of each amino acid (Table 3).

**Table 3:** The path in 20D space of the first 20 amino acids for the human ND6 protein, adapted from Randić [59]

n	AA	Coordinates	f	i	s	distance
1	M	0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0	1	1	1	1
2	M	0 0 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0	2	3	4	2
3	Y	0 1 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 0	1	1	5	2.236068
4	A	0 1 0 0 0 0 0 2 0 0 0 0 0 0 0 0 0 0 0 1	1	1	6	2.44949
5	L	0 1 0 0 0 0 0 2 0 1 0 0 0 0 0 0 0 0 0 1	1	1	7	2.645751
6	F	0 1 0 0 0 0 1 2 0 1 0 0 0 0 0 0 0 0 0 1	1	1	8	2.828427
7	L	0 1 0 0 0 0 1 2 0 2 0 0 0 0 0 0 0 0 0 1	2	3	11	3.316625
8	L	0 1 0 0 0 0 1 2 0 3 0 0 0 0 0 0 0 0 0 1	3	5	16	4
9	S	0 1 0 0 1 0 1 2 0 3 0 0 0 0 0 0 0 0 0 1	1	1	17	4.123106
10	V	1 1 0 0 1 0 1 2 0 3 0 0 0 0 0 0 0 0 0 1	1	1	18	4.242641
11	G	1 1 0 0 1 0 1 2 0 3 0 0 1 0 0 0 0 0 0 1	1	1	19	4.358899
12	L	1 1 0 0 1 0 1 2 0 4 0 0 1 0 0 0 0 0 0 1	4	7	26	5.09902
13	V	2 1 0 0 1 0 1 2 0 4 0 0 1 0 0 0 0 0 0 1	2	3	29	5.385165
14	M	2 1 0 0 1 0 1 3 0 4 0 0 1 0 0 0 0 0 0 1	3	5	34	5.830952
15	G	2 1 0 0 1 0 1 3 0 4 0 0 2 0 0 0 0 0 0 1	2	3	37	6.082763
16	F	2 1 0 0 1 0 2 3 0 4 0 0 2 0 0 0 0 0 0 1	2	3	40	6.324555
17	V	3 1 0 0 1 0 2 3 0 4 0 0 2 0 0 0 0 0 0 1	3	5	45	6.708204
18	G	3 1 0 0 1 0 2 3 0 4 0 0 3 0 0 0 0 0 0 1	3	5	50	7.071068
19	F	3 1 0 0 1 0 3 3 0 4 0 0 3 0 0 0 0 0 0 1	3	5	55	7.416198
20	S	3 1 0 0 2 0 3 3 0 4 0 0 3 0 0 0 0 0 0 1	2	3	58	7.615773

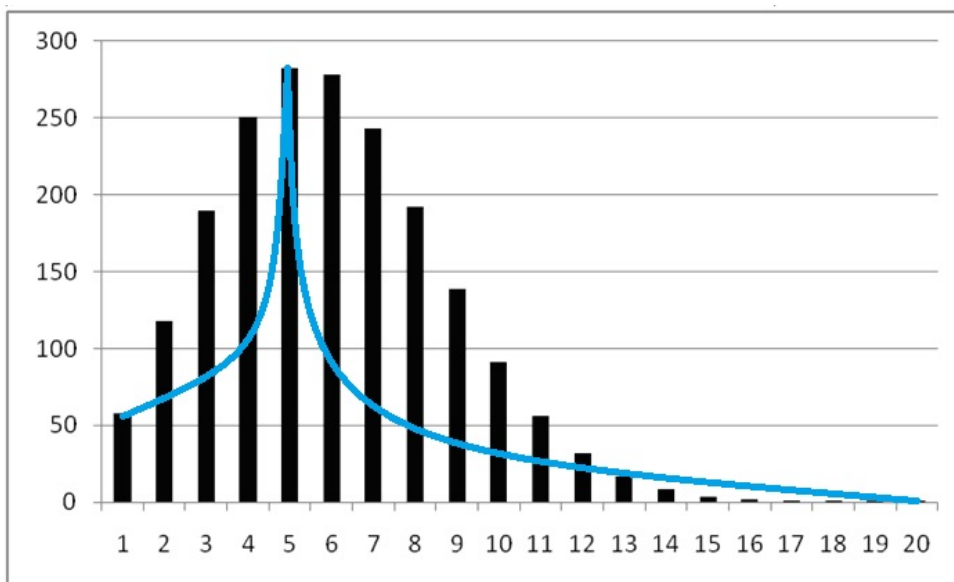
From these acquired eigenvalues, the increment number “i” can be used for obtaining the proteins profile [70] (Figure 8) according to Eq. (10).

$$\frac{\sum i^n}{n!} \quad (10)$$

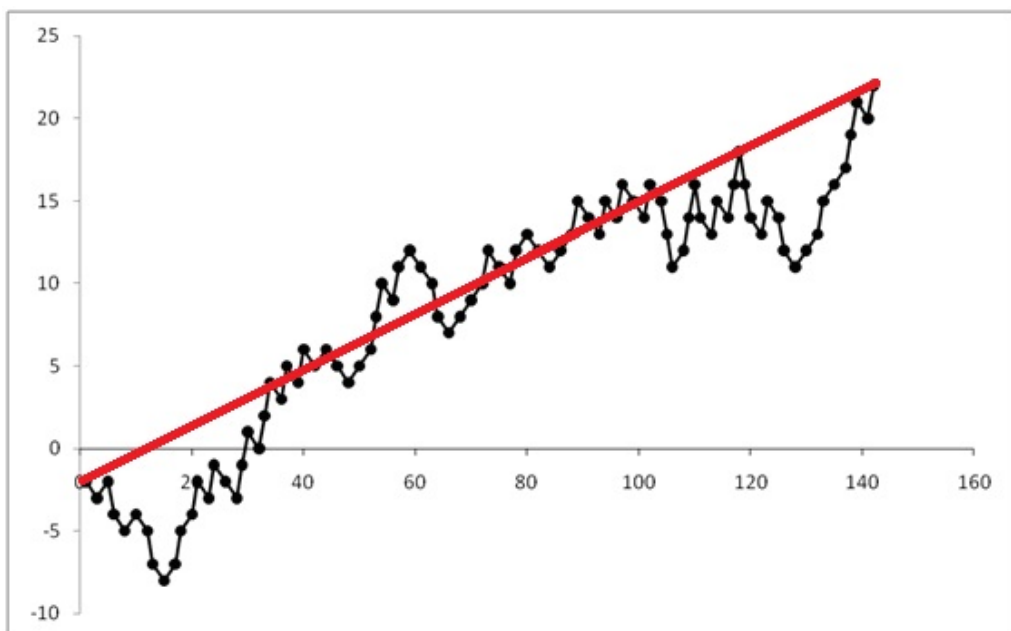
The protein profile offers visual comparison yet comes with loss of information. By assigning integer coordinates to the four nucleotides, we obtain a path which is the lattice representation of the sequence (Figure 9) [71].

The sequence can be represented as a 2D or 3D figure by using 2 (A:+1, -2; T:+2, -1; G:+2, +1; C:+1, +2) or 3 coordinates, however the 2D lattice representation is suited for rapid and easy comparison of genetic material [72]. The transition from nucleotides to amino acids is easily done by increasing the range of x, y coordinates, and ordered by their natural abundance [63] (Figure 10 and Figure 11). The lattice is a highly reduced and simple representation compared to the D/D matrix, but contains sufficient visible information for a fast preliminary comparison of genetic material; this is also achieved without loss of information. For a faster preliminary analysis, one can reduce the lattice only by counting by a set number of steps and skipping the rest, thus obtaining reduced zigzag curves [62].

**Figure 8:** Protein profile the first 20 amino acids for the human ND6 protein, adapted from Randić [59]; see text for details



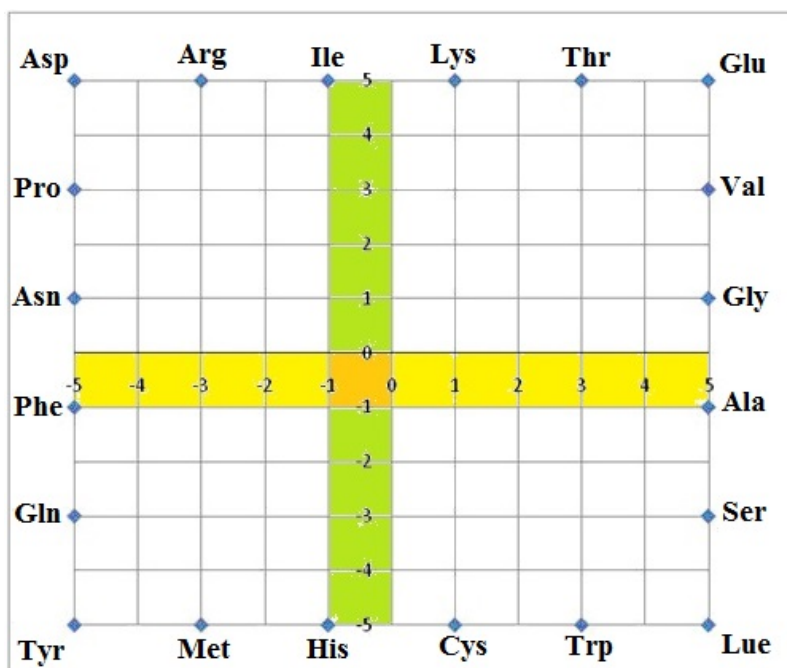
**Figure 9:** Lattice representation of the hypothetical RNA of the first 30 amino acids for the human ND6 protein, adapted from Randić [59]; see text for details



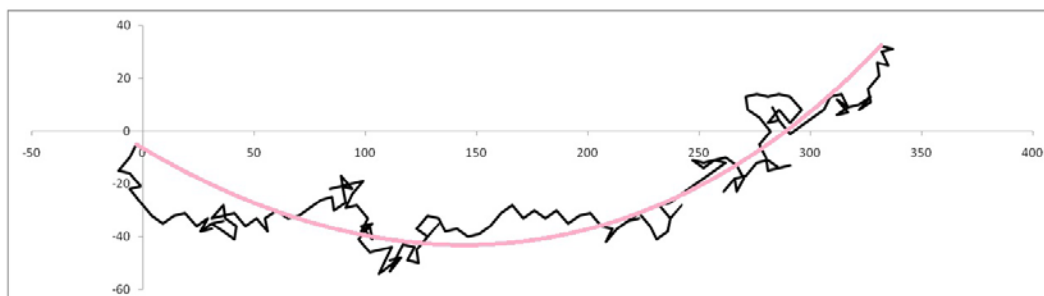
A different type of lattice representation involves constructing a 20X20 distance matrix from the average coordinates of each amino acid. Although this approach offers us a more

facile visual effect, it comes with the cost of information loss and requires supplementary processing [63].

**Figure 10:** Anticlockwise arrangement of the 20 amino acids in decreasing order of their abundance within proteins, adapted from Randić [63]; see text for details



**Figure 11:** Lattice representation for the human ND6 protein, adapted from Randić [59]; see text for details



Finally, worth mentioning the *proteomic maps* offering the necessary information on pathological processes and drug pathways; they can be obtained by separating proteins depending on their isoelectric charge and molecular mass, using electrophoresis and gel chromatography of cellular proteins as constant  $x$ ,  $y$  maps. The maps of the same organ cell differ only in the abundance of the proteins [73].

Proteomic maps are processed as bubble diagrams or as tables, for later reprocessing as:

- zigzag curves, by connecting proteins in order of their abundance [73];
- graph ordering, by constructing a mass to charge Cartesian system, in which all connections have a positive slope and there are no intermediate connections [74];
- graph clustering and assigning a third coordinate  $z$  as the size of the spot [75];
- nearest-neighbor adjacency graph in which each selected protein connects to its nearest  $n$  neighbors [76];
- Voronoi adjacency graph, which represents the area of points nearest to the protein [77].

#### 4. CONCLUSION

The importance of topological modeling of proteins is distinguished, as it presents simplified visual characteristics for the rapid and efficient modeling and comparison of protein structures. Even though graphical representations of proteins do not faithfully reproduce reality, they are a valuable tool for quick comparison of similarities and differences of genetic material.

As one would expect, in order to correctly identify the similarities of two biological sequences, a single representation or alignment is not enough and that is why it is necessary to compare and supplement one graphical representation or numerical characterization with another.

Current literature offers many possibilities for the graphical representation of genetic material, but it has not succeeded to provide an answer for the construction of DNA so far. That is why the aim of the authors is to pursue this query in future research.

#### REFERENCES

1. Alexanderson G.L. Euler and Königsberg's bridges: A historical view. *American Mathematical Society* **2006**, *43*, 567–573.
2. Harary F. *Graph Theory*; Addison-Wesley Publishing Company; 1969.
3. Bárány I. Applications of Graph and Hypergraph Theory in Geometry. *Combinatorial and Computational Geometry* **2005**, *52*, 31-50.
4. Erdős P.; Meir A.; Sós V.T.; Turán P. On some applications of graph theory, I. *Discrete Mathematics* **2006**, *306*, 853-866.
5. Heckmann T.; Schwanghart W.; Phillips J.D. Graph theory — recent developments of its application in geomorphology. *Geomorphology* **2014**.
6. Cvetković D. Applications of Graph Spectra: an Introduction to the Literature. *Zbornik Radova* **2009**, 7-32.
7. Balaban A.T. Applications of graph theory in chemistry. *Journal of Chemical Information and Computer Sciences* **1985**, *25*, 334-343.
8. Hamori E. Novel DNA sequence representations [7]. *Nature* **1985**, *314*, 585.
9. Randić M. Graphical representations of DNA as 2-D map. *Chemical Physics Letters* **2004**, *386*, 468-471.
10. Diestel R. *Graph theory*; Springer-Verlag; 2000.
11. Bender E.A.; Williamson S.G. *Lists, Decisions and Graphs*; University of California: San Diego; 2010.

12. Heilbronner E. Huckel molecular orbitals of Mobius-type conformations of annulenes. *Tetrahedron Letters* **1964**, 1923-1928.
13. Putz M.; De Corato M.; Benedek G.; Sedlar J.; Graovac A.; Ori O. Topological Invariants of Möbius-Like Graphenic Nanostructures. In *Topological Modelling of Nanostructures and Extended Systems*; Ashrafi A.R., Cataldo F., Iranmanesh A., Ori O., eds.; Springer Netherlands; 2013; 7 7, pp. 229-244.
14. Ajami D.; Oeckler O.; Simon A.; Herges R. Synthesis of a Mobius aromatic hydrocarbon. *Nature* **2003**, 426, 819-821.
15. Ivanciuc O.; Balaban A.T. *Graph Theory in Chemistry*; John Wiley & Sons; 1998.
16. Janezic D.; Lucic B.; Milicevic A.; Nikolic S.; Trinajstic N.; Vukicevic D. Hosoya Matrices as the Numerical Realization of Graphical Matrices and Derived Structural Descriptors. *Croatica Chemica Acta* **2007**, 80, 271-276.
17. Hosoya H. The Topological Index Z Before and After 1971. *Internet Electronic Journal of Molecular Design* **2002**, 1, 428-442.
18. Shirinivas S.G.; Vetrivel S.; Elango N.M. Applications of graph theory in computer science an overview. *International Journal of Engineering Science and Technology* **2010**, 2, 4610-4621.
19. Broder A.; Kumar R.; Maghoul F.; Raghavan P.; Rajagopalan S.; Stata R.; Tomkins A.; Wiener J. Graph structure in the Web. *Computer Networks: The International Journal of Computer and Telecommunications Networking* **2000**, 33, 309-320.
20. Sykes M.F.; Essam J.W.; Heap B.R.; Hiley B.J. Lattice constant systems and graph theory. *Journal of Mathematical Physics* **1966**, 7, 1557-1572.
21. Chui Y.H.; Rees R.J.; Snook I.K.; O'Malley B.; Russo S.P. Topological characterization of crystallization of gold nanoclusters. *Journal of Chemical Physics* **2006**, 125.
22. Brouwer D.H. Structure solution of network materials by solid-state NMR without knowledge of the crystallographic space group. *Solid State Nuclear Magnetic Resonance* **2013**, 51-52, 37-45.
23. Martineau C.; Senker J.; Taulelle F. NMR Crystallography. *Annual Reports on NMR Spectroscopy* **2014**, 82, 1-57.
24. Gutman I.; Trinajstić N. Graph theory and molecular orbitals. XV. The Hückel rule. *The Journal of Chemical Physics* **1976**, 64, 4921-4925.
25. Mu L.; Feng C. Topological research on molar diamagnetic susceptibilities for inorganic compounds. *MATCH Communications in Mathematical and in Computer Chemistry* **2007**, 58, 591-607.
26. Basak S.C.; Grumvald G.D.; Niemi G.J. Use of Graph-Theoretic and Geometrical Molecular Descriptors in Structure-Activity Relationships. In *From Chemical Topology to Three-Dimensional Geometry*; Balaban A.T., ed.; Kluwer Academic Publishers: Hingham, MA, USA; 1997, pp. 73-116.
27. Bonchev D.; Trinajstić N. Information theory, distance matrix, and molecular branching. *The Journal of Chemical Physics* **1977**, 67, 4517-4533.
28. Estrada E. Atom-bond connectivity and the energetic of branched alkanes. *Chemical Physics Letters* **2008**, 463, 422-425.
29. Elk S.B. Graph theoretical algorithm to canonically name the isomers of the regular polyhedranes. *Journal of Chemical Information and Computer Sciences* **1992**, 32, 14-22.
30. Balasubramanian K. Combinatorics of cluster enumeration. *Journal of Chemical Information and Computer Sciences* **1992**, 32, 47-54.

31. Rappoport D.; Galvin C.J.; Zubarev D.Y.; Aspuru-Guzik A. Complex Chemical Reaction Networks from Heuristics-Aided Quantum Chemistry. *Journal of Chemical Theory and Computation* **2014**, *10*, 897-907.
32. Randić M. Representation of molecular graphs by basic graphs. *Journal of Chemical Information and Computer Sciences* **1992**, *32*, 57-69.
33. Randić M.; Razineer M. On Characterization of 3D Molecular Structure. In *From Chemical Topology to Three-Dimensional Geometry*; Balaban A.T., ed.; Kluwer Academic Publishers: Hingham, MA, USA; 1997, pp. 157-236.
34. Skvortsova M.I.; Baskin I.I.; Stankevich I.V.; Palyulin V.A.; Zefirov N.S. Molecular Similarity. 1. Analytical Description of the Set of Graph Similarity Measures. *Journal of Chemical Information and Computer Sciences* **1998**, *38*, 785-790.
35. Ori O.; Cataldo F.; Graovac A. Topological Ranking of C28 Fullerenes Reactivity. *Fullerenes, Nanotubes and Carbon Nanostructures* **2009**, *17*, 308-323.
36. Iranmanesh A.; Ashrafi A.R.; Graovac A.; Cataldo F.; Ori O. Wiener Index Role in Topological Modeling of Hexagonal Systems – From Fullerenes to Graphene. In *Mathematical chemistry monographs*; Gutman I., Furtula B., eds.; University of Kragujevac and Faculty of Science Kragujevac: Kragujevac; 2012, pp. 135–155.
37. Vincent C.; Macleod B. An application of network theory to migration analysis. *Canadian Studies in Population* **1974**, *1*, 43.
38. McGuffin M.J.; Balakrishnan R. Interactive Visualization of Genealogical Graphs. Proceedings of *Proceedings of IEEE Symposium on Information Visualization (InfoVis 2005)*, Minneapolis, Minnesota, USA, Andrews K. ed., 2005, 17-24.
39. Kier L.B.; Hall L.H. Intermolecular Accessibility: The Meaning of Molecular Connectivity. *Journal of Chemical Information and Computer Sciences* **2000**, *40*, 792-795.
40. Herndon W.; Chen H.-T.; Zhang Y.; Rum G. QSAR Study of PAH Carcinogenic Activities: Test of a General Model for Molecular Similarity Analysis. In *Molecular Modeling and Prediction of Bioactivity*; Gundertofte K., Jørgensen F., eds.; Springer US; 2000, pp. 47-52.
41. Putz M.V.; Ori O.; Cataldo F.; Putz A.-M. Parabolic Reactivity “Coloring” Molecular Topology: Application to Carcinogenic PAHs. *Current Organic Chemistry* **2013**, *17*, 2816-2830.
42. Buzsáki G. *Rhythms of the Brain*; Oxford University Press; 2009.
43. Girvan M.; Newman M.E.J. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 7821-7826.
44. Barthélemy M. Betweenness centrality in large complex networks. *European Physical Journal B* **2004**, *38*, 163-168.
45. Salvador R.; Suckling J.; Coleman M.R.; Pickard J.D.; Menon D.; Bullmore E. Neurophysiological architecture of functional magnetic resonance images of human brain. *Cerebral Cortex* **2005**, *15*, 1332-2342.
46. Gong G.; He Y.; Concha L.; Lebel C.; Gross D.W.; Evans A.C.; Beaulieu C. Mapping anatomical connectivity patterns of human cerebral cortex using in vivo diffusion tensor imaging tractography. *Cereb Cortex* **2009**, *19*, 524-536.
47. Stam C.J.; Van Dijk B.W. Synchronization likelihood: An unbiased measure of generalized synchronization in multivariate data sets. *Physica D: Nonlinear Phenomena* **2002**, *163*, 236-251.
48. Ravasz E.; Barabási A.L. Hierarchical organization in complex networks. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics* **2003**, *67*, 261121-261127.



49. Amaral L.A.N.; Scala A.; Barthélemy M.; Stanley H.E. Classes of small-world networks. *Proceedings of the National Academy of Sciences of the United States of America* **2000**, *97*, 11149-11152.
50. Freeman L. A Set of Measures of Centrality Based on Betweenness. *Sociometry* **1977**, *40*, 35-41.
51. Carrillo-Reid L.; Hernandez-Lopez S.; Tapia D.; Galarraga E.; Bargas J. Dopaminergic modulation of the striatal microcircuit: receptor-specific configuration of cell assemblies. *J Neurosci* **2011**, *31*, 14972-14983.
52. Ray S.; Miller M.; Karalunas S.; Robertson C.; Grayson D.S.; Cary R.P.; Hawkey E.; Painter J.G.; Kriz D.; Fombonne E.; Nigg J.T.; Fair D.A. Structural and functional connectivity of the human brain in autism spectrum disorders and attention-deficit/hyperactivity disorder: A rich club-organization study. *Hum Brain Mapp* **2014**, *35*, 6032-6048.
53. Bassett D.S.; Bullmore E.; Verchinski B.A.; Mattay V.S.; Weinberger D.R.; Meyer-Lindenberg A. Hierarchical organization of human cortical networks in health and Schizophrenia. *Journal of Neuroscience* **2008**, *28*, 9239-9248.
54. Catani M.; Ffytche D.H. The rises and falls of disconnection syndromes. *Brain* **2005**, *128*, 2224-2239.
55. Bullmore E.; Sporns O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* **2009**, *10*, 186-198.
56. Randić M. 2-D graphical representation of proteins based on virtual genetic code. *SAR QSAR Environ Res* **2004**, *15*, 147-157.
57. Guo X.; Nandy A. Numerical characterization of DNA sequences in a 2-D graphical representation scheme of low degeneracy. *Chemical Physics Letters* **2003**, *369*, 361-366.
58. Randić M. Graphical representations of DNA as 2-D map. *Chemical Physics Letters* **2004**, *386*, 468-471.
59. Randić M.; Vračko M.; Plavšić D. Novel Graphical Representation of Proteins Based on the Virtual Genetic Code. *International Journal of Chemical Modeling* **2009**, *1*, 305-324.
60. Randić M.; Vračko M.; Nandy A.; Basak S.C. On 3-D graphical representation of DNA primary sequences and their numerical characterization. *Journal of Chemical Information and Computer Sciences* **2000**, *40*, 1235-1244.
61. Randić M.; Kleiner A.F.; DeAlba L.M. Distance/distance matrices. *Journal of Chemical Information and Computer Sciences* **1994**, *34*, 277-286.
62. Randić M.; Zupan J.; Balaban A.T.; Vikić-Topić D.; Plavšić D. Graphical representation of proteins. *Chem Rev* **2011**, *111*, 790-862.
63. Randić M.; Vračko M.; Nović M.; Plavšić D. 20. On uniform representation of proteins by distance matrix. In *Quantum Frontiers of Atoms and Molecules*; Putz M.V., ed.; Nova Science Publishers: New York; 2011, pp. 673.
64. Randić M.; Nović M.; Vračko M. On novel representation of proteins based on amino acid adjacency matrix. *SAR QSAR Environ Res* **2008**, *19*, 339-349.
65. Randić M.; Vračko M.; Lerš N.; Plavšić D. Novel 2-D graphical representation of DNA sequences and their numerical characterization. *Chemical Physics Letters* **2003**, *368*, 1-6.
66. Randić M.; Vračko M.; Lerš N.; Plavšić D. Analysis of similarity/dissimilarity of DNA sequences based on novel 2-D graphical representation. *Chemical Physics Letters* **2003**, *371*, 202-207.

67. Liao B.; Chen W.; Sun X.; Zhu W. A binary coding method of RNA secondary structure and its application. *Journal of Computational Chemistry* **2009**, *30*, 2205-2212.
68. Jeffrey H.J. Chaos game representation of gene structure. *Nucleic Acids Research* **1990**, *18*, 2163-2170.
69. Barnsley M.F. Chapter IX Measures on Fractals In *Fractals Everywhere*, 2nd ed ed.; Academic Press: Boston, MA; 1988.
70. Nović M.; Randić M. Representation of proteins as walks in 20-D space. *SAR QSAR Environ Res* **2008**, *19*, 317-337.
71. Hamori E.; Ruskin J. H curves, a novel method of representation of nucleotide series especially suited for long DNA sequences. *Journal of Biological Chemistry* **1983**, *258*, 1318-1327.
72. Nandy A. A new graphical representation and analysis of DNA sequence structure: I. Methodology and application to globin genes. *Current Science* **1994**, *66*, 309.
73. Randić M. On Graphical and Numerical Characterization of Proteomics Maps. *Journal of Chemical Information and Computer Sciences* **2001**, *41*, 1330-1338.
74. Randić M.; Zupan J.; Novič M. On 3-D Graphical Representation of Proteomics Maps and Their Numerical Characterization. *Journal of Chemical Information and Computer Sciences* **2001**, *41*, 1339-1344.
75. Bajzer Ž.; Randić M.; Plavšić D.; Basak S.C. Novel map descriptors for characterization of toxic effects in proteomics maps. *Journal of Molecular Graphics and Modelling* **2003**, *22*, 1-9.
76. Randić M.; Lerš N.; Plavšić D.; Basak S.C. On invariants of a 2-D proteome map derived from neighborhood graphs. *Journal of Proteome Research* **2004**, *3*, 778-785.
77. Randić M.; Novič M.; Vračko M.; Plavšić D. Study of proteome maps using partial ordering. *Journal of Theoretical Biology* **2010**, *266*, 21-28.