On the Characterization of Molecular Stereostructure: 1. Cis−Trans Isomerism

Giorgi Lekishvili

Department of Chemistry, I. Javakhishvili Tbilisi State University, 380028, 3, Chavchavadze avenue,
Tbilisi, Georgia

The definition of the graph isomorphism has been generalized for a more precise description of molecular structure. The adjacency matrix and different topological indices were modified with the help of the complex numbers. Some of the compounds were characterized. The obtained indices were substantiated to be more isomer sensitive than the 2D ones. The quantitative structure−property relationship was ascertained in an alkene series ($r > 0.99$).

INTRODUCTION

Graph theory is widely used for modeling various chemical and biological processes. Its main purpose is to ascertain the generalities of the dependence of different properties of compounds on the structure of their molecules. Based on the discovery and the description of quantitative structure−property (QSPR) and the quantitative structure−activity (QSAR) relationships, one can carry out the given class compounds systematization, investigation of their transformations, the computer-assisted planning of synthesis, and the prognosis of features of hypothetical substances.[1−5] The above mentioned generalities enable one to systemize available information. The methods based on topological indices do not require the "difficult mathematics". The computer techniques development and many available applying programs have conditioned further increase of the methods popularity. The QSAR and the QSPR studies accuracy is determined by the precision of characterizing the chemical structure with the topological indices. The definition of the chemical structure involves not only the atoms connectivity[1] but also their dislocation in space. Many simple 2D (two dimensional) indices do not consider the information about the heteroatoms and the stereostructural features that restricts the field and the efficiency of their application. Recently Randić has proposed a new way of solving the first problem via considering any matrix as a "weighted" graph.[5−7] The main obstacle for characterization of the steric features is inherent limitation of the chemical graph theory, which is devoid of utilizing 3D information on a structure. Presently there exist more than 100 descriptors, among which one can find Randić, Kier−Hall, Balaban, Hosoya, and Bonchev−Trinajstić indices, etc. Therefore, it might have been less important to introduce new ones, as to modify the old descriptors being still of some interest.

SEVERAL DEFINITIONS OF THE GRAPH THEORY

Pursuant to the graph theory,[8] a molecule can be presented with a graph. Graphs $G = \{V,E\}$ are discrete algebraic objects comprising two finite, countable sets $V = \{v_i | i = 1,2,...,n\}$ (its elements are named vertices) and $E = \{e_k = \{v_{ki},v_{kj}\}_k | k = 1,2,...,m\}$ (the elements, edges, represent nonordered (generally) pairs of the elements of V. The edges can be multiple.[10] The elements of a graph can differ with "weights". Atomic numbers, electronegativities, bond polarities, etc. can be employed as weights in the chemical context.[2,10]

Each graph can be characterized with topological indices, which are its invariants. The graph invariants are numbers, which have the identical meanings for every graph isomorphic of the given one. Two graphs G_1 and G_2 are isomorphic $G_1 \cong G_2$, if a bijection exists between their V_1 and V_2 sets preserving the adjacency. As one can easily note, the molecular graphs corresponding to conformers and stereoisomers are isomorphic, so they are described with the identical numbers (invariants). The most important representation of a graph is the adjacency matrix $A = \{a_{kl}\}$, is where
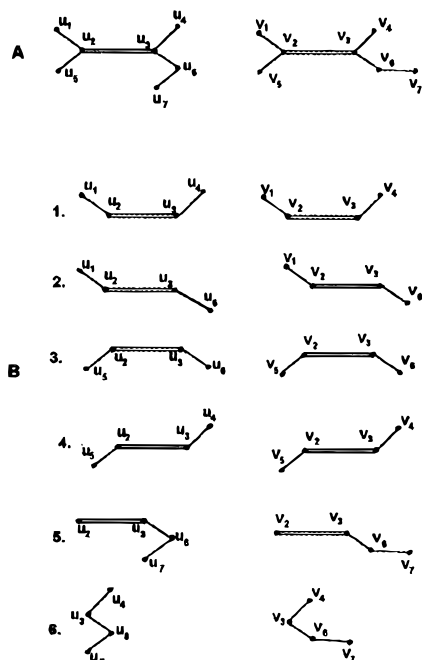
$$a_{kl} = \begin{cases} 1, & \text{if } \{k,l\} \in E, \text{ i.e., if the } k\text{th and } l\text{th} \\ & \text{vertices are connected} \\ 0, & \text{otherwise}^8 \end{cases}$$

It is evident that modification of the definition of graph isomorphism is the crucial problem.[11] The present work deals with the mathematical description of stereoisomers and conformers without any additional datum of the quantum chemistry or other of the similar type.

MODIFICATION OF THE ISOMORPHISM DEFINITION

The two conformers of 2,3-dimethylpent-2-ene shown in Figure 1 are depicted with the different but isomorphic graphs. Consequently they have the identical topological indices. In order to consider the spatial features of the conformers, all of the nonbranched subgraphs with the path length $d = 3$ have been detached from the molecular graphs. Within the suggested approach, these subgraphs are called the basic subgraphs (BSG). Two vertices k and l connected with a BSG are named the extreme vertices; the BSG is called the own basic subgraphs (OBSG) of the kth and lth vertices, while its central edge is defined as the own central edge (OCE). From the point of view of location of the extreme vertices, a BSG can consist four forms only (Figure 2, see also ref 12). As it is evident from Figure 2, at least one vertex is collinear with the corresponding OCE in the **b**, **c** subgraphs. The extreme vertices are on the same side

---

MOLECULAR STEREOSTRUCTURE. 1. CIS—TRANS ISOMERISM

J. Chem. Inf. Comput. Sci., Vol. 37, No. 5, 1997 **925**

Figure 1. (A) The graphs corresponding to two conformers of 2,3-dimethyl-pent-2-ene and (B) their basic subgraphs. The numeration of the vertices is arbitrary.
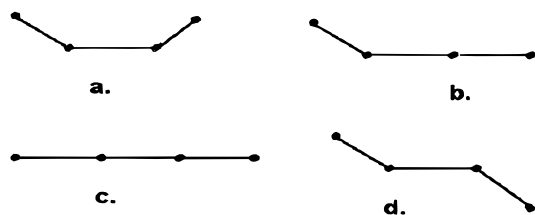

Figure 2. The forms of a basic subgraph (the isomers of tetracene can be represented using these subgraphs, see also ref 12).

toward their OCE in the **a** subgraphs, while they hold the opposite positions in the **d** BSG.

Let us create the pairs of the vertices connected with at least one BSG, provided none of them (vertices) are collinear with the OCE. The set of these pairs will be marked with the capital letter Q. The Q set can be divided into two subsets Q_c and Q_t for the pairs of the vertices being, accordingly, in the cis and trans positions toward their OCE-s.

On the base of the above said, the definition of isomorphism can be extended to stereographs. Here two graphs G_1 and G_2 are isomorphic $G_1 \cong G_2$ only when there exists bijection between their V_1 and V_2 sets preserving the adjacency and the identity of the location of the extreme vertices of all of the BSG-s toward their OCE-es. This approach can be considered to be a logical development of the chemical graph theory in compliance with the Trach—Zefirov model.[13] Evidently the two different graphs depicted in Figure 1 cannot be considered isomorphic anymore, because the mapping of the u_2, u_3, u_6, u_7 vertices accordingly on the v_2 v_3, v_6, v_7 ones infringes the identity of the location of the extreme vertices toward their OCE (Figure 1, fifth BSG). The same situation is observed when mapping the u_4, u_3, u_6, u_7 vertices on the v_4, v_3, v_6, v_7 ones correspondingly (Figure 1, sixth BSG).

The most important representation of graphs is the adjacency matrix. In order to describe the spatial features of graphs the complex numbers were used. The elements
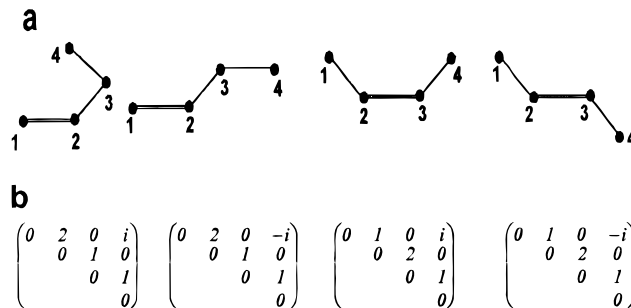
Figure 3. (a) The isomeric butenes with the conformers and (b) the modified adjacency matrices.

of the matrix[11,14] are defined as follows

$$a_{kl} = \alpha_{kl} + i\beta_{kl} \tag{1}$$

where

$$\alpha_{kl} = \begin{cases} \mu_{kl}, & \text{if } \{k,l\} \in E, \text{ i.e., if the } k\text{th and } l\text{th} \\ & \text{vertices are connected} \\ 0, & \text{otherwise} \end{cases} \tag{1a}$$

μ_{kl} is the bond order. Usually $\mu = 1; 2; 3; 3/2$, for single, double, triple, and aromatic bonds, respectively

$$\beta_{kl} = \begin{cases} 1, & \text{if } \{k,l\} \in Q_c, \text{ i.e., if } k\text{th and } l\text{th atoms} \\ & \text{connected with a BSG are on the} \\ & \text{same side toward their OCE} \\ -1, & \text{if } \{k,l\} \in Q_t, \text{ i.e., if } k\text{th and } l\text{th atoms} \\ & \text{connected with a BSG are on the} \\ & \text{opposite sides toward their OCE} \\ 0, & \text{otherwise} \end{cases} \tag{1b}$$

In rings where k and l vertices can be connected with nBSG ($n \geq 1$), the following formula has been suggested (similarly, the path lengths can be calculated in cycles[15]):

$$\beta_{kl} = \frac{1}{n}\sum_j \beta_{klj} \tag{1c}$$

As it is clear, **the meanings of β_{kl} do not depend on the order of vertex labels.**

Evidently, if the introduced model is successful, the different conformers and isomers must differ with their modified adjacency matrices. Two conformers of butene and *cis*- and *trans*-but-2-ene and their graphs and matrices are shown in Figure 3.

As one can see, their matrices are different. The approach enables one to describe stereoisomers and conformers via modifying the topological indices based on the adjacency matrix.

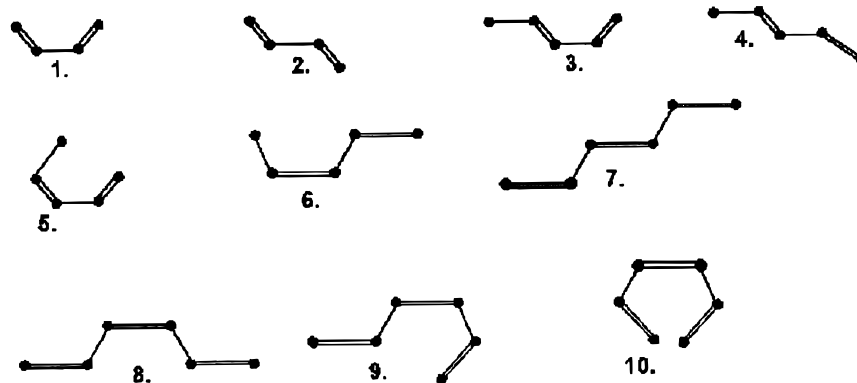Generally, all of the modified indices of this kind are of the following form:

$$\text{TI} = A + iB \tag{2}$$

The topological indices (TI) used in ascertaining the QSP and QSA relationships must be the real numbers (TI $\in$ R). Furthermore different models based on the eq 2 can be used. For example

$$\gamma_{\text{TI}} = \sqrt{A^2 + nB^2} - mB|B| \tag{3}$$

where n and m parameters could be varied for acquiring

Figure 4. Several dienes, trienes, and their conformers The graphs are embedded in a plane.

Table 1. Some of the Isomeric Dienes and Trienes with the Corresponding γ_w Indices (see Figure 4)

no.	γ_w	no.	γ_w
1	9.9499	6	20.0000
2	10.1499	7	36.0283
3	20.0000	8	35.1143
4	20.4996	9	34.9193
5	19.6996	10	34.2283

better correlations, when $\beta = 0$, $\gamma_{TI} = A$. The two refined and most widespread indices: the Randić index[16] and the Wiener index[17] have been modified to characterize the molecular structure.

FEATURES OF THE MODIFIED WIENER INDEX

The Wiener index suggested in 1947 can be calculated as[17]

$$W = \frac{1}{2}\sum_k\sum_l d_{kl} \qquad (4)$$

where d_{kl} is the path length between the kth and lth vertices. In the outlined approach d_{kl} has been changed as follows

$$\sigma_{kl} = d_{kl} + i\beta_{kl} \qquad (5)$$

Accordingly, the Wiener index was modified as

$$W^{(i)} = \frac{1}{2}\sum_k\sum_l \sigma_{kl} \qquad (6)$$

In order to estimate its efficiency, we used the model based on eq 3

$$\gamma_w = \sqrt{A^2 + B^2} - 0.1B|B| \qquad (7)$$

i.e., $n = 1$ and $m = 0.1$ have been selected.

The different conformers of some dienes and trienes depicted in Figure 4 were examined. The boat and the chair conforms of cyclohexane were also considered (γ_w(chair) = 45.1111, γ_w(boat) = 44.6444) The results are shown in Table 1. As the obtained data reveal, one can successfully characterize stereospecifities of molecules by using γ indices. On the basis of the results we conclude that the new model is sensitive to geometry variations of stereoisomers. Hence, γ_w can be considered a 3D-descriptor.

It is important to note that the γ_w index reflects the generally observed property of other 3D indices. If a conformer is more compact, the magnitude of the index is

smaller. This fact demonstrates the suggested model to be logical, since the Wiener index is constructed on the base of the distance matrix. It is also interesting that the valuable 3D descriptors introduced earlier by Randić and Trinajstić (3D-ID)[18] and by Estrada (Ω^c)[19] differing from the γ_w index decrease when the compactness of conformers increases. This fact substantiates the plentifulness of joint studies of QSPR and QSAR by using different models.

APPLICATIONS OF THE MODIFIED RANDIC INDEX

The molecular connectivity index was introduced by Randić. It can be calculated as follows[16]

$$\chi = \sum_j [(\delta_k \delta_l)^{-1/2}]_j \qquad (8)$$

where d_k is the degree of the kth vertex (summation is carried out over all edges):

$$\delta_k = \sum_l a_{kl} \qquad (9)$$

(a_{kl} are the elements of the unmodified adjacency matrix).

The Randić index found a wide application in solving different problems of mathematical chemistry.

The further modification of the index was suggested by Kier and Hall in order to extend its use to heteroatoms and molecules having the bond multiplicity[20]

$$\delta_k^v = Z_k^v - h_k \qquad (9a)$$

Here Z_k^v is the number of valence electrons of the kth atom and h_k is the number of the hydrogen atoms bonded with the kth atom.

Within the given approach, the vertex degree (its absolute meaning is in general, not integer, but real number) is defined as

$$\delta^{(i)}_k = \sum_l a_{kl} \qquad (10)$$

where a_{kl} are taken from the eq 1. Hence, the Randić index has now the following form:

$$\chi^{(i)} = \sum_j [(\delta^{(i)}_k \delta^{(i)}_l)^{-1/2}]_j \qquad (11)$$

In order to demonstrate the use of the modified Randić index and the model based on eq 3 we have carried out the special investigations despite the fact that the purpose of this

MOLECULAR STEREOSTRUCTURE. 1. CIS—TRANS ISOMERISM

J. Chem. Inf. Comput. Sci., Vol. 37, No. 5, 1997 **927**

Table 2. Alkenes with Their Entropies, γ_χ and χ^v Indices

no.	alkenes	S^0_{298}	χ^v	γ_χ
1	ethene	55.45	0.5000	0.5000
2	propene	63.80	0.9856	0.9856
3	butene	73.04	1.5236	1.5236
4	*cis*-butene	71.90	1.4881	1.4025
5	*trans*-butene	70.86	1.4881	1.3749
6	2-Me-propene	70.17	1.3536	1.3536
7	pent-1-ene	82.65	2.0235	2.0235
8	*cis*-pent-2-ene	82.76	2.0260	1.9750
9	*trans*-pent-2-ene	81.36	2.0260	1.9363
10	2-Me-butene	81.73	1.9143	1.9143
11	2-Me-but-2-ene	80.92	1.8661	1.8661
12	3-Me-butene	79.70	1.8963	1.8963
13	hex-1-ene	91.93	2.5235	2.5235
14	*cis*-hex-3-ene	90.73	2.5639	2.5536
15	*trans*-hex-3-ene	89.59	2.5639	2.5066
16	2-Me-pent-1-ene	91.34	2.4143	2.4143
17	3-Me-pent-1-ene	90.06	2.4342	2.4342
18	4-Me-pent-1-ene	87.89	2.3794	2.3794
19	2-Me-pent-2-ene	90.45	2.4040	2.4040
20	*cis*-4-Me-pent-2-ene	89.23	2.3988	2.2882
21	*trans*-4-Me-pent-2-ene	88.02	2.3988	2.2685
22	2-Et-butene	90.01	2.4750	2.4750
23	2,3-diMe-butene	87.39	2.2971	2.2971
24	2,3-diMe-butene	82.16	2.1969	2.1969
25	2,3-diMe-but-2-ene	86.67	2.2500	2.2500

paper was not concerned with applications to the QSPR. The following model (based on the eq 3) was used

$$\gamma_x = \sqrt{A^2 + 3B^2} - 0.1B|B| \qquad (12)$$

(here $n = 3$; $m = 0.1$).

We selected 25 alkenes C2—C6 for the study. Experimental entropies[21] S^0_{298}[20] vs χ^v and γ_x indices are given in Table 2.

As usual, the linear correlation between the topological index and the property (S^0_{298}) was searched.

$$S^0_{298} = a_1\gamma_x + b_1 \qquad (10a)$$

$$S^0_{298} = a_2\chi^v + b_2 \qquad (10b)$$

Here $a_1 = 17.81$, $b_1 = 46.56$, $r_1 = 0.9925$, $s_1 = 1.19$, and $F_1 = 1523$ and $a_2 = 17.79$, $b_2 = 46.13$, $r_2 = 0.9919$, $s_2 = 1.24$, and $F_2 = 1402$.

The obtained results demonstrate that the new approach has enabled us to achieve the better correlation.

ON THE MODELS BASED ON THE COMPLEX TOPOLOGICAL INDICES

The outlined approach has important flexibility that is not characteristic for most traditional topological indices. The novel descriptors γ_w and γ_χ have free parameters n and m which can be optimized for individual applications. Hence, γ_w and γ_χ belong to what has recently been described as fourth generation molecular descriptors.[22] Moreover, topological indeces γ_i seem to be the first fourth generation 3D molecular descriptors.

One may begin the search for optimal descriptors in QSPR or QSAR with the initial conditions $n = 1$; $m = 0.1$. If the correlation obtained is better than when $n = m = 0$ (i.e., the correlation based on the corresponding topological index), this will indicate how to vary the parameters. As soon as we obtained a better result in our study, we gave up further

search for optimal correlation (it was not within the scope of our research) as our purpose was to demonstrate the suitableness of the model.

The possibility of varying the parameters is a desirable feature to introduce new indices which are based on the old ones in order to curb proliferation of descriptors.[23] We believe also that one should prefer the models of that kind rather than generalizing the approach by using analytical functions of complex variables.

CONCLUDING REMARKS

The presented modification of the definition of isomorphism enables one to characterize the molecular spatial features. The introduced approach makes it possible to discover and to describe the QSP and the QSA relationships, because many properties of compounds are directly or indirectly determined with the stereostructural characteristics (molecular size, symmetry, chirality, etc.). Using complex numbers, the Wiener and Randić indices have been modified. As the statistical parameters of correlation elucidated, more precise results were obtained by applying the novel models. Evidently, the outlined approach enables one to modify also the other indices based both on the distance and on the adjacency matrices (e.g., Balaban's *J* index). The presence of the free parameters "*n*" and "*m*" makes it possible to achieve better correlations in individual applications. The heteroatoms contained in a molecule can be also considered via a correct choice of tabulated parameters as weights. Consequently, one has to continue with further research to explore the new resources in the combinatorics and the graph theory. That is why the further development and the generalization of the chemical graph theory is of importance, particularly, with respect to extensions to 3D structures. The present work is a contribution to solve this problem.

Further investigations and elaboration of the new topological indices on the basis of the suggested approach are in progress.[23] Also, it appears promising to codify molecules via the use of the modifications of the adjacency matrix by complex numbers.[24]

REFERENCES AND NOTES

(1) Balaban, A. T. From Chemical Graphs to 3D Molecular Modeling. In *From Chemical Topology to Three Dimensional Geometry*; Balaban, A. T., Ed.; Plenum Press: New York and London, 1996.

(2) Stankevich, M. I.; Stankevich, I. V.; Zefirov, N. S. Topologicheskiye Indeksy v Organicheskoi Khimii (Topological Indices in Organic Chemistry). *Usp. Khim.* **1988**, LVII, 3, 337—366 (Russian).

(3) Trinajstić, N. Chemical Graph Theory, 2nd revised ed.; CRC: Boca Raton, FL, 1992.

(4) Balaban, A. T. *Chemical Application of Graph Theory*; Academic: London, 1976.

(5) Randić, M. The Nature of Chemical Structure. *J. Math. Chem.* **1990**, *4*, 157—184.

(6) Estrada, E. Edge Adjacency Relationship in Molecular Graphs Containing Heteroatoms: A New Topological Index Related to Molar Volume. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 4, 701—707.

(7) Lekishvili, T.; Lekishvili, G.; Alexidze, N. Mathematical Investigation of the Cancerogenic Activity of Some of the Hormones. *Bull. Acad. Sci. Georg.* **1996** (in press).

(8) Harary, F. Graph Theory; Edition-Wesley Publishing Company: CA, 1969. The Russian translation: *Teoriya Grafov. Mir*: Moscow, 1973; pp 15—179.

(9) Caufmann, A. Introduction to the Combinatories from the Point of View of Applications. Dunod: Paris, 1968.

(10) Gverdtsiteli, M.; Gamsiani, G. Organuli Naertebis Chatseris Algebruli Metodi. (Algebraic Method of the Codification of Organical Compounds). TSU *Shromebi*. **1987**, *256*, 41—44 (Georgian).

(11) Lekishvili, G. Alkenebis Cis-trans Isomeriuli Gardakmnebis Matematikuri Gamokvleva (Mathematical Investigation of Cis Irans Isomeric Transformations of Alkenes). *Proc. Acad. Sci. Georg.* **1994**, *20*, 1−4 (Georgian).

(12) Balaban, A. T.; Harary, F. Chemical Graphs 5. Enumeration and Proposed Nomenclature of Benzenoid Cata-condensed Polycyclic Aromatic Hydrocarbons. *Tetrahedron* **1968**, *24*, 2505−2516.

(13) Trach, S. S.; Zefirov, N. S. Principy Simmetrii i Sistemnosti v Khimii (The Principles of Symmetry and Systemity in Chemistry). Stepanov N. F., Ed.; MGU Press: Moscow, 1987; pp 54−86 (Russian).

(14) Lekishvili, G. Algebraicheskaya Kharakteristika Stereoisomerov (Algebraical Characterization of Stereoisomers). *International Conference on Chemistry*; dedicated to 75 Year Anniversary of I. Javakhishvili TSU. Collection of Abstracts; Tbilisi, 1994; p 108.

(15) Lekishvili, G. A Parametric Model for Construction of the Topological Indices of Information Type. *Bull. Acad. Sci. Georg.* In press.

(16) Randić, M. On characterization of Molecular Branching. *J. Am Chem. Soc.* **1975**, *69*, 6609−6615.

(17) Wiener, H. J. Structural Determination of Paraffin Boiling Points. *J. Am. Chem. Soc.* **1947**, *69*, 17−20.

(18) Randić, M. Jerman-Blazic, B.; Trinajstić, N. Development of 3 Dimensional Molecular Descriptors. *Comput. Chem.* **1990**, *14*, 237−246.

(19) Estrada, E. Three- Dimensional Molecular Descriptors Based on Electron Charge Density Weighted Graphs. *J. Chem Inf. Comput. Sci.* **1995**, *35*, 4, 708−713.

(20) Kier, L. B.; Hall, L. B. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.

(21) Kireev, V. A. Metody Practicheskikh Raschotov v Termodinamike Khimicheskhikh Reaktsii (Methods of Practical Calculations in Thermodynamics of Chemical Reactions). Moscow, Khimiya, 1975; pp 462−463 (Russian).

(22) Randić, M. Linear Combinations of Path Numbers as Molecular Descriptors. *New J. Chem.* Submitted for publication.

(23) Lekishvili, G. On Characterization of Molecular Stereostructure. 2. New 3- Dimensional Distance-Based Indices. Work in progress.

(24) Lekishvili, G. On Characterization of Molecular Stereostructure. 3. A New Way for Codification of Molecules. Work in progress.