match no. 26 pp. 215-228 1991

ON THE CORRELATION BETWEEN THE MOLECULAR INFORMATION TOPOLOGICAL AND MASS-SPECTRA INDICES OF ORGANOMETALLIC COMPOUNDS

Skorobogatov V.A., Konstantinova E.V.
Institute of Mathematics Siberian Division of the USSR,
Academy of Sciences, 630090 Novosibirsk 90, USSR

Nekrasov Yu.S., Sukharev Yu.N., Tepfer E.E.
A.N. Nesmeyanov Institute of Organoelement Compounds, USSR,
Academy of Sciences, 28, Vavilov Str., 117334, Moscow , USSR

(received: July 1990)

ANNOTATION

Mass-spectrum of a molecule contains information on the ways of molecule fragmentation and on the probability of each fragment to be formed as a result of intramolecular decay occuring upon electron impact, therefore it can be used for estimating reactive capability of a molecule.

At present, no theoretical model is available that would permit to evaluate the reactive capability of a compound when given its structural formula. To solve this problem one can use topological indices that are associated with the mass-spectrum characteristics.

In the present work some information topological indices based on the metric characteristics of molecular graphs and therefore reflecting specific structural features of a molecule are proposed along with information indices of mass-spectra which characterize the mass-spectral properties of molecules and their reactive capabilities under the conditions of electron impact fragmentation. The correlation of molecular and spectral indices is studied; for a set of thirty five ferrocene derivatives $C_pFeC_5H_4R$, it is established that the initial set

of compounds is divided into three subsets by linear regression. In the cases considered the correlation factors varied from 0.80 to 0.97.

INTRODUCTION

One of the key problems of modern chemistry is to establish a relation between the structure and reactive capability of molecules. The traditional approach to this problem consists of the search for quantitative characteristics of processes (kinetic and thermodynamic parameters of reactions) as functions of electron and structural parameters of molecules [1]. Generalized characteristics describing both the structure and reactive capability of molecules are of the greatest interest. Introduction of generalized characteristics allows to order the whole set of compounds according to their chemical activity with respect to molecular substituents and special structural features of the molecular skeleton. The compact form of presentation of the molecular structure and reactivity can be used for predicting the reaction types for molecules of a given class; for the search of new compounds within the given range of reactive capability characteristics; for the classification and identification of compounds or for the search of structural analogs; and for the selection of reagents when solving the problems of synthesis. At present, the topological indices of molecular graphs are widely used as structural invariants [2].

The choice of generalized indices of reactive capability is a more complicated problem since, as a rule, total schematics of liquid phase reactions including the data on either outputs or rate constants for all the final and intermediate products are not known. To this end, one can use the mass - spectrum as the mapping of the series - parallel process of intramolecular dissociation of molecular ions, because it contains full information on the probability of appearance of each product formed upon fragmentation of the studied compound under electron impact. As the mass-spectrum generalized characteristic one can use the information index that is

calculated by the formula of Shannon [3].

The fact that there really exists a correlation between mass-spectral characteristics and reactive capability of a molecule is supported by the correlation between relative intensities of ion mass-spectrum peaks and either substituent constants or activation energies of corresponding processes for ion intramolecular dissociation [3].

On the whole, mass-spectra reconstruct an approximate picture for the structure and reactive capability of molecules by displaying the "behavior" of some of the molecular fragments, which can be interpreted as subgraphs of molecular graphs. Topological indices [2,6], in particular, metric characteristics [5] allow also an approximation which describes the molecular graphs as a whole, evaluates the graph compactness (or branching), and describes the position of individual subgraphs (fragments) in the molecular graph.

For predicting the molecular properties on the basis of spectral information, the conventional integral topological indices have some disadvantages since they are not structurized and do not provide local characterization for individual molecular fragments (including those corresponding to the mass-spectrum ions).

The generalized indices, which take into account the spectral information features, can be obtained on the basis of new topological characteristics - graph topological spectra [7] which provide the local characterization of individual molecular fragments. In order to estimate the relation between topological characteristics and mass-spectrum characteristics an information approach is utilized [6,9] providing a method to quantitatively analyze various ways of presenting chemical structures and compare these ways in terms of the same quantitative scale.

TOPOLOGICAL SPECTRA OF MOLECULAR GRAPHS

Let G(V,E) be a finite non-oriented connected non-marked graph without loops and multiple edges. The graph G corresponding to molecular structure will be called the

molecular graph. A fixed set of subgraphs of the molecular graph G is denoted as $H = \{H_i | i=1...k\}$.

Let the pair $\langle G,H \rangle$ be associated to the structurized topological index which is in fact a family F of sets of topological characteristics $F^j(H)=\{f^j(H_1),f^j(H_2),...,f^j(H_k)\}$. Further we shall denote $f^j(H_i)=f_i^j$ and $F=\{f_i^j | i=1...k,\ j=1...n\}$.

Definition. A topological n-dimensional spectrum $S_n(G)$ of a graph G is a collection of values of sets of topological characteristics $F^j(H)$, $j=1...n$, when mapping $F \longrightarrow E^n$ of a family of topological characteristics to Euclidean space.

A topological two-dimensional spectrum (n=2) will be called simply a topological spectrum and a topological two-dimensional spectrum of a single-vertex subgraph, i.e. $H_i \simeq \langle v_i \rangle$, will be called a topological vertex spectrum.

In the case where metric characteristics are considered as topological ones, the spectrum will be called the metric spectrum.

Fig.1 shows an example of the metric vertex spectrum $S_2(G)$ for a tree. The set $H=\{v_i | i=1...5\}$ is represented by vertices of a tree G and functions f^1 and f^2 are metric characteristics: $m_2(v)$ is the vertex mean square deviation and $e(v)$ is the vertex eccentricity, respectively [5].

For plotting a two-dimensional spectrum let us draw its fractions $[(f^1(v_i),f^2(v_i));(f^1(v_i),0)]$ and call these the spectral lines. In the case where $f^1(v_i)=f^1(v_j)$ assume that $f^2(v_i,v_j)= f^2(v_i)+f^2(v_j)$ and the corresponding line is a 2-multiple one. If there are k coinciding values of function f^1, the spectral line is k-multiple.

The spectral line corresponding to vertices 4 and 5, given in Fig.1, is multiple and $e(v_4,v_5)=e(v_4)+e(v_5)=6$.

Let us give one more example of the metric vertex spectrum construction. To this end, let us consider some notions from Ref.[5].

The layer matrix of a graph $G=(V,E)$, $|V(G)|=p$, is called a matrix $\lambda(G)=\|\lambda_{ij}\|$, $i=1...p$, $j=1...d(G)$, where λ_{ij} is equal to the number of vertices located at a distance j from vertex i, d(G) is a diameter of graph G. The matrix l(G) consisting of

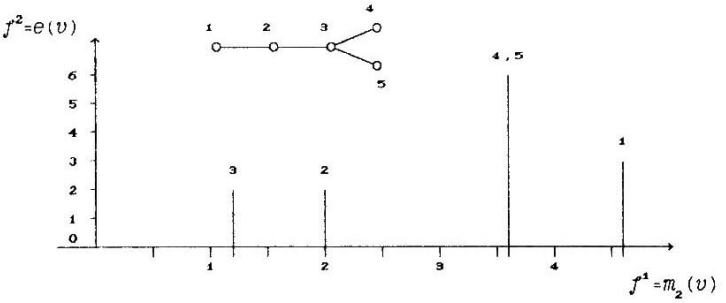


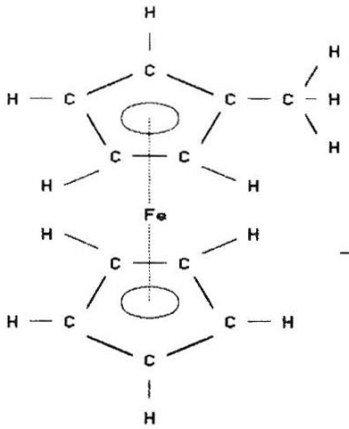
Fig.1. Metric vertex spectrum $S_2(G)$.

all mutually different (if arranged in pairs) rows of the layer matrix is called 1-spectrum. The graphs G_1 and G_2 are called isometric, if there can be established a one-to-one correspondence on the sets of their vertices so that distance is conserved. As is known from Ref.[5], graphs G_1 and G_2 are isometric $G_1 <\sim> G_2$, if their corresponding 1-spectra are the same, i.e. $l(G_1)=l(G_2)$. The isometricity ratio $G_1 <\sim> G_2$ divides the vertex sets of these graphs into the isometricity classes.

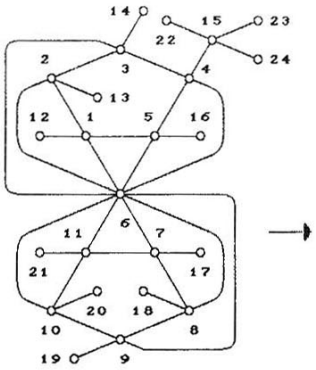
Fig.2 (a,b,c) shows the layer matrix with separated classes of autometricity for the molecular graph of methylferrocene . The matrix is canonical: its lines are length decrease ordered and then the lines of the same length are ordered lexicographically. The numeration of autometricity classes X_i, i=1...8, corresponds to the line numeration in the canonical layer matrix.

The metric vertex spectrum of graph G on the basis of autometricity ratio can be defined for functions f^1 and f^2 such that f^1 corresponds to the graph autometricity classes and f^2 is the number of vertices in each class. The multiplicity of lines of such a spectrum will coincide with f^2 .

Fig.2(d) shows the described metric vertex spectrum for the methylferrocene molecular graph.



a) the structure formula of methylferrocene

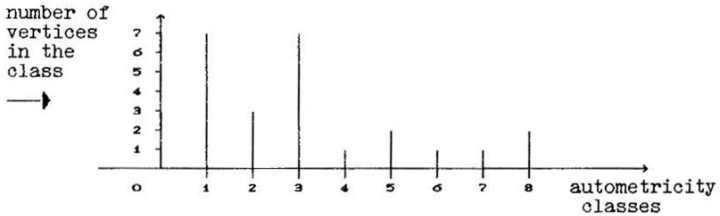


b) the molecular graph of methylferrocene

$\lambda(G)=$

1	3	9	7	3	12,13,17,18,19,20,21	$= X_1$
1	3	3	9	7	22,23,24	$= X_2$
4	9	7	3		7, 8, 9,10,11, 1, 2	$= X_3$
4	3	9	7		15	$= X_4$
1	3	9	10		16,14	$= X_5$
10	10	3			6	$= X_6$
4	12	7			4	$= X_7$
4	9	10			3, 5	$= X_8$

c) the canonical layer matrix and autometricity classes of the methylferrocene molecular graph



d) The vertex metric spectrum of the methylferrocene molecular graph

Figure 2.

The examples considered above are topological vertex spectra where metric characteristics were used as their functions. The use of metric characteristics of fragments (subgraphs) of the graph [5] as f^j characteristics enables one to get metric fragmentary spectra.

The degenerate case of topological n-dimensional spectrum is a 1-dimensional spectrum for the pair <G,H>, where H={G}, whose set of topological characteristics consists of a single element, being the characteristic of the graph G. Such characteristics are ,for example, following metric characteristics: the eccentricity of graph e(G), the distance of graph D(G), the average distance of graph $D_{av}(G)$ etc., as well as any other integral characteristics of graphs.

Thus, here are considered some examples of a new class of structurized topological characteristics of graphs. The use of these characteristics for construction of information theoretic topological indices [6], will be illustrated further for the problem of a structure-reactive capability for organometallic compounds.

INFORMATION INDICES OF METRIC VERTEX SPECTRA AND MASS-SPECTRA

The following well-known principle [6] is generally used for the construction of information indices.

Let X be a set consisting of n elements.Let us assume that by some equivalence criterium the elements are divided into N equivalence classes X_i, so that $n = \sum_{i=1}^{N} n_i$ where n_i is the number of elements in subset X_i. Then $p_i = n_i/n$ is the probability for a single element to belong to the i-th subset, and to estimate quantitatively the information that corresponds to one element of the set one can use the distribution entropy for the set elements defined by the following formula of Shannon [3]:

$$H = -\sum_{i=1}^{N} p_i . log_2 p_i \qquad (1)$$

For construction of information indices for molecular

structures in theoretical chemistry atoms and chemical bonds are regarded as the set elements; information on the molecular symmetry is also used.[1]

Information indices of molecular graphs were constructed for various matrices of graphs (adjacency matrix, cycle matrix) and also for some topological indices such as the Wiener, Hosoya and Randić indices , the Balaban centric index.[2]

In the work presented here, correlations between the generalized spectrum characteristics and the molecular structural invariants are studied. Information indices of the mass-spectrum calculated according to formula (1) in Ref.[8] were used to estimate of reactive capability of molecules.

Two indices were considered:

1) HA information index of spectrum with no account of ion masses, i.e.,

$$p_i = A_i \Big/ \sum_{j=1}^{N} A_j \qquad (2)$$

2) HS information index of spectrum, taking into account the ion masses, i.e.,

$$p_i = \frac{m_i \cdot A_i}{\sum_{j=1}^{N} m_j \cdot A_j} \qquad (3)$$

where A_i is an amplitude (intensity) of the mass-spectrum i-th peak, m_i is the ion mass corresponding to the i-th peak, N is the number of peaks in the mass-spectrum.

The information indices of metric vertex spectra of molecular graphs were considered as structural invariants of molecules.

Information metric index H_1 of the metric spectrum described above, based on the autometricity ratio,is defined by

---

[1] References to original works can be found in [2,6].

[2] See previous footnote.

the following formula:

$$H_1 = -\sum_{i=1}^{N} \frac{f_i^2}{n} \cdot log_2 \frac{f_i^2}{n} \qquad (4),$$

where f_i^2 is the number of vertices in the i-th class of autometricity, and $n = \sum_{i=1}^{N} f_i^2$ is the total number of graph vertices.

The information metric index H_2 of the metric vertex spectrum, for which the vertex distance of graph D(v) was used as a function f^1, with f^2 being equal to the number of vertices with the same distances, coincides with that described in Ref.[6]. The index is calculated by the formula:

$$H_2 = -\sum_{i=1}^{N} \frac{D(v_i) \cdot n_i}{2D(G)} \cdot log_2 \frac{D(v_i) \cdot n_i}{2D(G)} \qquad (5),$$

where D(G) is the graph distance, n_i is the number of vertices of distance $D(v_i)$.

Also considered was the information index H_3 defined for the degenerate metric spectrum with $f^1 = D_{av}(G)$. It is calculated by the following formula:

$$H_3 = log_2 D_{av}(G) \qquad (6)$$

In this case, since the spectrum is given by the value of the metric characteristic $D_{av}(G)$ and there is no splitting into equivalence classes , the formula of Shannon cannot be used.

For the given indices a linear regression analysis was carried out in order to determine correlations between these indices and the information indices of mass-spectra.

DISCUSSION OF RESULTS

A group of mass-spectra of ferrocene derivatives $C_5FeC_5H_4R$, where R is a series of substituents given in Table 1, was chosen as a subject of the study. The calculated HA indices ordered in succession of their decreasing values, HS indices and corresponding values of calculated values H_1, H_2, H_3 for all 35 substituents are also given in Table 1. Fig 3 shows some examples of HA as a function of H_1 for three sets from Table 2. HA as a function of H_2 and H_3 , and HS as a function of H_1, H_2, H_3 can be recovered by Tables 1 and 2.

An analysis of the results has shown that the whole set of constituents can be split into three sets, each of them belonging to its corresponding regression, and for these cases, the correlation ratio ranges from 0.80 to 0.97.

Table 1. Calculated values of indices HA, HS, H_1, H_2, H_3 for ferrocene derivatives.

N	R	HA	HS	H_1	H_2	H_3
1	OCOPh	2.5839	2.5319	9.5786	9.5690	7.2429
2	H	2.6041	2.2099	1.2286	1.1323	5.6845
3	CN	2.9443	2.5489	2.4911	2.4484	5.8237
4	CH=CHCN-trans	3.1147	2.7550	3.0349	3.0588	6.3508
5	$CH=CH_2$	3.2164	2.7319	2.8917	2.8023	6.2110
6	COOH	3.2958	2.9859	2.7807	2.7780	6.0870
7	OCOMe	3.4096	2.9304	2.9750	2.9306	6.5142
8	$CMe=CH_2$	3.4769	2.8748	3.0196	2.9627	6.5146
9	CHO	3.6745	3.3840	2.5539	2.4952	5.9440
10	COPh	3.6840	2.9998	3.4898	3.4727	7.0615
11	Me	3.7455	3.4297	2.5826	2.5031	6.0498
12	CH=CHPh	3.8222	2.9978	3.6695	3.6418	7.2790
13	COOMe	3.8757	3.4056	2.9750	2.9219	6.4954
14	COEt	4.0759	3.6152	3.1918	3.0744	6.6518
15	$C(Me)(OH)(CH_2CN)$	4.0847	3.7701	3.4957	3.4011	6.9186
16	COMe	4.0876	3.5524	2.8521	2.7923	6.3192
17	CH_2OH	4.0909	3.6768	2.8317	2.8108	6.1785
18	NEt_2	4.2017	3.7428	3.1061	2.9810	7.0923
19	CH=CHCOCH=CHPh	4.2370	3.5820	4.1063	4.0663	7.8470
20	I	4.3749	3.8098	1.2266	1.1329	5.6848
21	CH(OH)Me	4.3807	4.1483	3.0910	3.0469	6.4884
22	$CH(OH)CMe_3$	4.3879	4.0019	3.1395	2.9822	7.1925
23	CH(OOCMe)Me	4.4591	4.1325	3.4440	3.3841	7.0027
24	CH=CHCOMe	4.4989	3.9226	3.9050	3.2671	6.8210
25	CH_2Ph	4.8016	4.3056	3.5197	3.4989	7.1011
26	C(Me)=CHMe	4.8441	4.4327	3.2506	3.1865	6.8074
27	CH=CHCOPh	4.9994	4.0742	3.8229	3.7932	7.4792
28	COCH=CHPh	5.0484	4.2742	3.8229	3.7950	7.4814
29	$COCF_3$	5.1005	4.8241	2.8521	2.7923	6.3192
30	$CH=C_5H_4$	5.2698	4.6130	3.2993	3.2608	6.8755
31	$CH(OH)CH_2Ph$	5.3283	4.9568	3.8452	3.8127	7.3995
32	C(Ph)=CHCOOH	5.3980	4.6880	3.8979	3.9159	7.3792
33	$(CH_2)_4COOH$	5.3770	4.6810	3.7448	3.7394	7.2699
34	CSPh	5.5972	4.9702	3.4898	3.4727	7.0615
35	$CH=C(CONH_2)_2$	6.2301	6.0075	3.9079	3.2214	7.0654

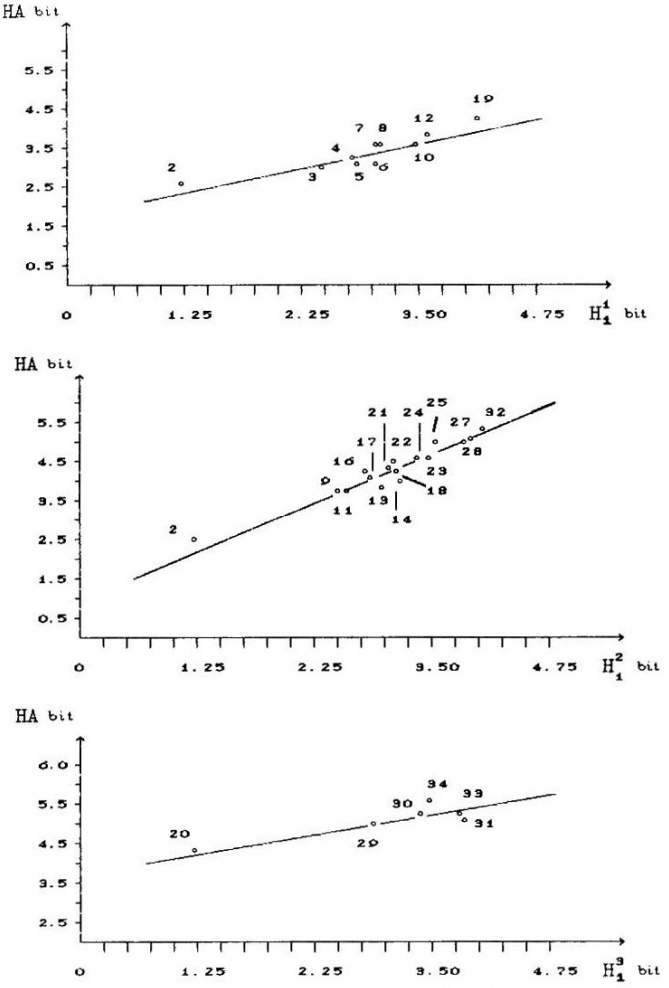


Figure 3. The dependence of HA on H_1^1, H_1^2, H_1^3 for the first three samplings of Table 2.

Table 2. Correlation coefficients r for all subsets of substituents.[3]

indices			substituent numbers from Tables 1	r
HA	H_1	H_1^1	2, 3, 4, 5, 6, 7, 8,10,12,19	0.94
		H_1^2	2, 9,11,13,14,16,17,18,21,22, 23,24,29,24,25,27,28,32	0.97
		H_1^3	20,29,30,31,33,34	0.94
	H_2	H_2^1	3, 4, 5, 6, 7, 8,10,12,19	0.95
		H_2^2	2, 9,11,13,14,16,17,18,21,22, 23,24,25,27,28,31,32	0.97
		H_2^3	20,29,30,31,33,34	0.95
	H_3	H_3^1	2, 3, 4, 5, 6, 7, 8,10,12,19	0.97
		H_3^2	9,11,13,14,15,16,17,18,21,22, 23,24,25,27,28	0.87
		H_3^3	20,29,30,31,32,33,32	0.88
HS	H_1	H_1^1	1, 3, 4, 5, 6, 7, 8, 9,10,15	0.92
		H_1^2	11,12,13,14,16,17,18,19,21,22, 23,24,25,26,27	0.81
		H_1^3	20,28,29,30,31,32,33,34	0.91
	H_2	H_2^1	1, 3, 4, 5, 6, 7, 8, 9,10,15	0.91
		H_2^2	11,12,13,14,16,17,18,19,21,22, 23,24,25,26,27	0.80
		H_2^3	20,28,29,30,31,32,33,34	0.91
	H_3	H_3^1	1, 3, 4, 5, 6, 7, 9,10,15	0.94
		H_3^2	11,12,13,14,16,17,18,19,21,22, 23,24,26,27	0.86
		H_3^3	20,25,28,29,30,31,33,34	0.92

[3] Values H_i^j, i=1...3, j=1...3, are taken from Table 1 according to the number of the substituent in the given Table.

This fact shows that some of substituents belonging to the same subset have some common properties which could be stipulated by the nature and number of atoms in the subset as well as by the substituent steric factors. The index linear dependence can be considered as a criterion of adequacy for models representing the generalized characteristics of spectrum and structure.

CONCLUSION

In conclusion, it is worth noting that the structure-topological analysis of molecular graphs and calculations of metric characteristics , based on the analysis, were carried out with no regard either for the atom's nature or molecular bond types. Accounting for the weight factors in molecular graphs, study of the ion element composition and neutral loss observed during fragmentation, will provide the correction of the studied correlations between the mass-spectral and structural indices taking into account their partitionings on the base of mass-spectral experiments. In addition, the partition, observed in spaces (HA,H_i), i=1...3, and (HS_i,H), i=1...3, of the whole set of substituents into its subsets within the same class of compounds indicates an imperfection of selected information metric indices. Their use should be regarded merely as an attempt to determine the dependence between indices of different nature but reflecting one and the the same physical essences of information sources. In this connection, the necessity arises to develop indices which could refer the compounds with different substituents to the same class and which they would have a selectivity for various classes of compounds. In this direction it seems very promising to use information indices constructed on the basis of topological fragmentary spectra for providing the characterization of certain fragments of a molecule and taking into account the features of spectral information.

REFERENCES

1. Zhdanov Yu.A., Minkin V.I. Correlation Analysis in Organic Chemistry. Rostov on Don. Edited Rostov on Don University. 1966.

2. Stankevich M.I., Stankevich I.V., Zefirov N.S. Topological Indices in Organic Chemistry //Uspeckhy khim. 1988. V.57, N.3. P.337-366.

3. Shannon C.E., Weaver W. Mathematical Theory of Communications. University of Illinois, Urbana, 1949.

4. Structure and Reactivity of Organic Compounds Ions in Gaseous Phase.//Ed. Tolsticov G.A. Ufa. 1986. 147 P.

5. Skorobogatov V.A., Dobrynin A.A. Metric Analisys of Graphs // Math. Chem. (MATCH). 1988. N.23. P.105-151.

6.Bonchev D. Information Indices for Atoms and Molecules.// Math. Chem. (MATCH). 1979. N.7. P.65-113.

7. Mzhelskaya E.V., Skorobogatov V.A. Metric Spectra of Molecular Graphs //Mathematical questions of the chemical information. Novosibirsk. N.130. Computer Systems P.68-83. 1990.

8.Robertson D.H. and Reed R.I. Proc. 19th Annu. Conf. Mass Spectrom. Allied Top., p.68 (1971).

9. Nekrasov Yu.S., Sukharev Yu.N., Molgacheva N.S., Tepfer E.E., Zagorevskii D.V., Skorobogatov V.A., Mzhelskaya E.V. Information Indices of Mass-Spectra and their Correlation with Invariants of Molecular Graphs of Organometallic Compounds //Heads of Reports to VIII All-Union Conference on Computers Use in Spectroscopy of Molecules and Chemical Investigations. Novosibirsk, 1989. P.286.