Establishment of correlations between the structure and reactivity of molecules in the gas phase based on information theory

Yu. S. Nekrasov, Yu. N. Sukharev, E. E. Tepfer,* and N. S. Molgacheva

A. N. Nesmeyanov Institute of Organoelement Compounds, Russian Academy of Sciences,
28 ul. Vavilova, 117813 Moscow, Russian Federation.
Fax: 007 (095) 135 5085. E-mail: tpfr@ineos.ac.ru

A general approach to revealing correlations between the structure of molecules and their reactivity in fragmentation processes under electron impact conditions based on the use of generalized structural and mass spectral characteristics is suggested. The characteristics were obtained using information theory, molecular graphs, and absolute reaction rates. Information topological indices of molecular graphs were used as generalized structural characteristics of molecules. They are a quantitative measure of the structural complexity of molecules and are expressed in information units. The gas-phase process of fragmentation of molecules under electron impact was used as a general reaction series for all volatiles. In terms of information theory, the mass spectrum represents the distribution of probabilities of the formation of ions of each type, and the information entropy of this distribution appears to be an integral characteristic of the reactivity of a molecule during fragmentation under electron impact in the gas phase. Using organic and organometallic compounds of several classes (ferrocene derivatives, arylsilanes, aromatic azo compounds, *etc.*) as examples, linear correlations between the information indices of the mass spectra and the information topological indices of the appropriate molecular graphs or electronic parameters of molecules have been found, which testifies that the approach suggested is adequate.

Key words: mass spectrometry, reactivity; ferrocene derivatives, arylsilanes, aromatic compounds; correlation equations, information indices.

The establishment of interrelations between the structure of molecules and their reactivity is one of the key problems of modern chemistry. Two main questions appear in the solution of this problem. The first of them is to find a reaction that is common to all compounds studied, and the second is to choose the parameters of the chemical process and the structural parameters of the substance that correspond to each other.

In the chemistry of solutions, it is almost impossible to find one common reaction occurring under identical conditions (medium, temperature, and others) to compare the reactivity of chemical substances of different classes. In addition, the majority of the reactions in solutions are complicated by solvation effects and, as a consequence, the measured parameters of the reactivity (rate constants, equilibrium constants, activation energies, and heats of reactions) are assigned not to the inherent (intrinsic) properties of the molecules or ions involved, but to complex ion-molecular systems. For example, a comparison of the heats of nucleophilic substitution

$$CH_3Br + Cl^- \rightarrow CH_3Cl + Br^-$$

in the gas (-10.9 kcal mol^{-1}) and liquid (15.7 kcal mol^{-1}) phases shows that the energy of desolvation of the transition state in solution (26.6 kcal mol^{-1}) is substantially greater than the heat of the reaction,[1] *i.e.*, the energy parameters of the reaction measured in solution reflect the solvation process to a greater extent than the nucleophilic substitution itself.

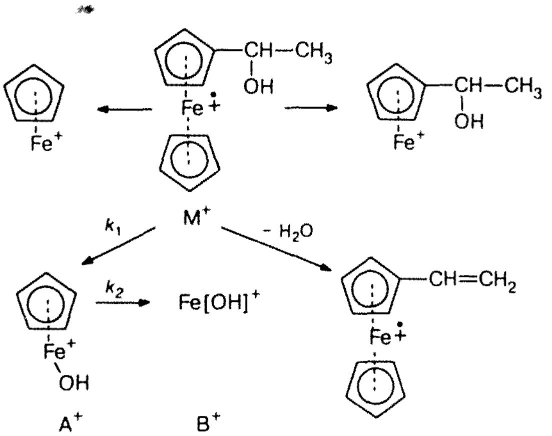
By contrast, gas-phase processes exclude the effect of the solvent. Thus, the gas-phase acidity has been determined for more than a thousand organic and organometallic compounds from the data of high-pressure mass spectrometry and ion cyclotron resonance.[2] However, even in this case, one can estimate the reactivity of only a limited range of molecules and under the conditions of only one, albeit quite common, acid-base interaction.

Mass spectral fragmentation of molecules can be proposed as a common process for all volatiles. Two experimentally determined parameters of a mass spectrum contain all necessary information on the reactivity of individual molecules: the mass numbers and the amplitudes of the peaks of ions appearing due to the sequential-parallel processes of intramolecular decomposition of a compound under electron impact. In this case, the qualitative characteristics of the reactivity are given by the types of reactions of formation or decomposition of ions. In particular, two types of processes are typical of the fragmentation of ferrocenylmethylcarbinol:

Translated from *Izvestiya Akademii Nauk. Seriya Khimicheskaya,* No. 11, pp. 2683—2688, November, 1996.

reactions involving simple bond cleavage and bond rearrangement (Scheme 1).

Scheme 1



$$M^+ \xrightarrow{k_1} A^+ \xrightarrow{k_2} B^+$$

$$\frac{dA}{dt} = k_1[M] - k_2[A] = 0$$

$$\frac{[A^+]}{[M^+]} = \frac{k_1}{k_2}$$

Reactions involving the elimination of π-bound ligands belong to the first type. The products of rearrangement reactions, ions and eliminated neutral fragments, contain new bonds that were absent in the initial molecule. Processes involving dehydration and the migration of a hydroxy group to a metal atom also belong to this type.[3]

A quantitative measure of the degree of fragmentation *via* a given route is the ratio of the intensities of daughter and parent ions, which, according to the formal kinetics laws for monomolecular stationary processes, is equal to the ratio of the rate constants of the formation and decomposition reactions for a given ion[4] (Scheme 1). Therefore, one may treat the relative intensities of mass spectral peaks in the same manner as reaction rate constants. Based on this, the traditional approach to the problem of correlations between the structure of a molecule and its reactivity in the gas phase, as in the chemistry of solutions, would be to search for dependences of the quantitative parameters of fragmentation processes on the electronic and structural parameters of molecules.

In terms of this approach we have established that, in particular, the degrees of fragmentation of diphenylmercury and alkylmercury halide derivatives, which involves cleavage of the mercury—carbon bond, correlate linearly with the Brown and Taft σ constants of the substituents, respectively (Scheme 2).[5]

Scheme 2

$$\underset{M^+}{[RC_6H_4\text{-}Hg\text{-}C_6H_4R]^{\dot{+}}} \xrightarrow{-RC_6H_4} \underset{A^+}{RC_6H_4Hg^+}$$

$$\log\frac{[A^+]}{[M^+]} = -0.26 + 0.56\sigma^+ \; (r = 0.984;\ n = 10)$$

$$\underset{M^+}{[R\text{-}Hg\text{-}Br]^+} \longrightarrow \underset{A^+}{HgBr^+}$$

$$\log\frac{[A^+]}{[M^+]} = -0.37 - 4.10\sigma^+ \; (r = 0.983;\ n = 7)$$

In the case of benzenechromiumtricarbonyl derivatives, the degree of dissociation of the metal—ligand bond correlates linearly with the number of vibrational degrees of freedom of the decomposing ion[6] (Scheme 3), and in the case of nonsubstituted metallocenes, the degree of dissociation of the metal—Cp bond depends linearly on the frequencies of the vibrations of the metal—ligand bond[3] (Scheme 4).

Scheme 3

$$\underset{M^+}{[RC_6H_5Cr(CO)_3]^{\dot{+}}} \xrightarrow{-3CO} \underset{B^+}{RC_6H_5Cr^+} \longrightarrow \underset{A^+}{Cr^+}$$

$$\log\frac{[A^+]}{[M^+]} = 1.59 - 0.022N \; (r = 0.981;\ n = 23)$$

Scheme 4

$$\underset{M^+}{Cp_2M^{\dot{+}}} \longrightarrow \underset{A^+}{CpM^+}$$

$$\log\frac{[A^+]}{[M^+]} = 9.90 - 4.16\log\nu \; (r = 0.973;\ n = 7)$$

All regularities established are analogous to isoenergetic, isoentropy, and isostructural reaction series, which are well known in the chemistry of solutions, and they are in entire correspondence with the theory of absolute reaction rates, which gives the following simplified expression for the rate constant of monomolecular decomposition of individual particles that are not in thermodynamic equilibrium with the medium:[7]

$$\log k = \log\nu - S\log(1 - \varepsilon/E), \qquad (1)$$

where ν is the frequency or the entropy factor, which is taken to be equal to the frequency of the vibration of the cleaved bond for processes of simple bond cleavage, ε is the activation energy, E is the intrinsic energy of the ion, and S is the number of effective oscillators, which is equal to $(3N - 6)/3$ (3 is the empirical coefficient) for a molecule consisting of N atoms.

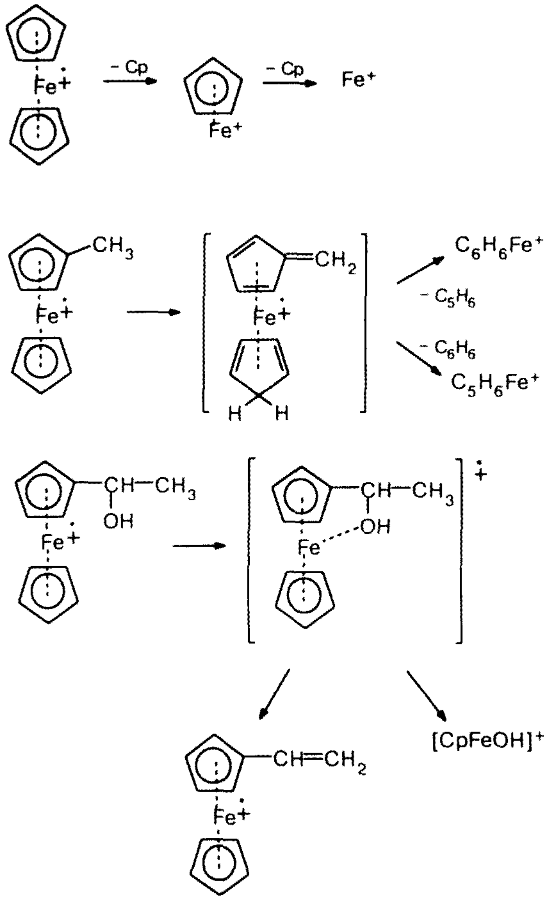
Hence it follows that mass spectra contain rich information about the reactivity of molecules and ions in the gas phase. The scheme of fragmentation gives an idea of the types of processes of intramolecular dissociation into ions, and the intensities of the peaks give information about the extent of the occurrence of these processes. In addition, studying the mass spectra of a series of related compounds shows the main factors determining the reactivity of ions in the gas phase. For example, the fragmentation of organomercury compounds is determined only by electronic effects of substituents in the ligands, while the size of a substituent becomes the main factor in benzenechromiumtricarbonyl derivatives, and the entropy factor is predominant in metallocenyl cations.

Thus, the study of the fragmentation of molecules under electron impact in the gas phase makes it possible to substantially broaden the concept of reaction series compared to that in the chemistry of solutions, because there is no solvent in this case and all processes occur under standard conditions. Using this approach, which copies many features of the methodology of studies of liquid-phase reactions, in terms of a single reaction series, one can quantitatively compare the reactivity of a limited range of molecules having a common direction of fragmentation.

In all examples considered, the main process is cleavage of the metal—ligand bond, while the contributions of other routes of decomposition to the fragmentation scheme are insignificant, which corresponds to the idea of the single reaction series. However, this situation is the exception rather than the rule. In most cases, even small changes in the structure of a molecule result in a qualitative change in its reactivity. Let us consider, *e.g.*, the fragmentation of three structurally related compounds of the ferrocene series (Scheme 5).

Ferrocene itself decomposes according to a scheme of sequential elimination of Cp ligands. Migration of the hydrogen atom of the methyl group to the nonsubstituted ring followed by the elimination of cyclopentadienyl and fulvene molecules is typical of methylferrocene. In the case of ferrocenylcarbinol, dehydration and migration of the hydroxy group to the metal atom are predominant.[3] The quantitative comparison of the reactivity of these molecules cannot be performed in terms of the approach considered above, which requires the conservation of a common mechanism for all compounds studied and the use of differential, *i.e.*, related to one common process, kinetic parameters. Therefore, we suggest another approach based on the search for common parameters of the reactivity of molecules in the gas phase, which are formed on the basis of the information theory using not one, as in the first case, but the whole set of processes of fragmentation of the compound studied.

Scheme 5

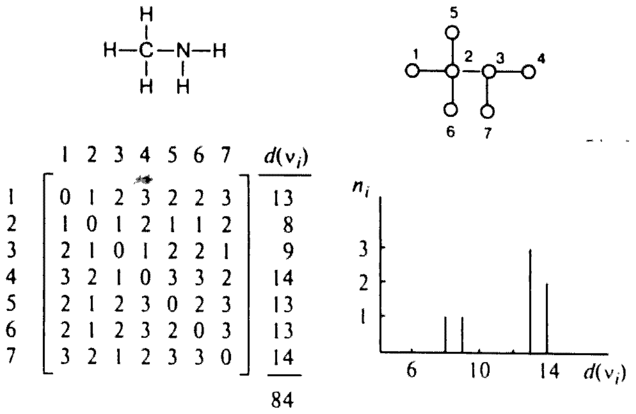


From the point of view of information theory, a mass spectrum is the distribution of probabilities of formation of ions of each kind (p_i), which are calculated as the ratio of the intensity of a given ion (I_i) to the total intensity of all ions of the spectrum ($p_i = I_i/\Sigma I_j$). Substituting the values of these probabilities into the Shannon formula (2), we obtain the amplitude information index H_{MS} (3), which is the integral characteristic of the reactivity of the molecule studied under the conditions of its fragmentation under electron impact in the gas phase.[8]

$$H = -\Sigma p_i \log_2 p_i, \qquad (2)$$

$$H_{MS} = -\Sigma(I_i/\Sigma I_j)\log_2(I_i/\Sigma I_j). \qquad (3)$$

Information topological indices of molecular graphs are widely used presently as an integral structural parameter for establishing structure—property correlations.[9—11] The structure of a molecule of N atoms, *e.g.*, methylamine, is replaced by its mathematical image, a molecular graph in which each atom corresponds to a vertex and each chemical bond corresponds to an edge. The molecular graph can be presented in matrix form, for example, as a matrix of distances in which the value of the length of the shortest chain of bonds connecting two vertices is put at the



$$\begin{array}{c|ccccccc|c}
 & 1 & 2 & 3 & 4 & 5 & 6 & 7 & d(v_i) \\
\hline
1 & 0 & 1 & 2 & 3 & 2 & 2 & 3 & 13 \\
2 & 1 & 0 & 1 & 2 & 1 & 1 & 2 & 8 \\
3 & 2 & 1 & 0 & 1 & 2 & 2 & 1 & 9 \\
4 & 3 & 2 & 1 & 0 & 3 & 3 & 2 & 14 \\
5 & 2 & 1 & 2 & 3 & 0 & 2 & 3 & 13 \\
6 & 2 & 1 & 2 & 3 & 2 & 0 & 3 & 13 \\
7 & 3 & 2 & 1 & 2 & 3 & 3 & 0 & 14 \\
\hline
 & & & & & & & & 84
\end{array}$$

$p_i = n_i/N$

$H_{MG} = -\Sigma(n_i/N)\log_2(n_i/N)$

Fig. 1. Formation of information topological indices of molecular graphs.

intersection of the numbers of these vertices. Summing these values in a row (or column) of the matrix, we obtain distance parameters of vertices of the given molecular graph. Using this parameter, the whole set of N vertices can be divided into classes of equivalence (n_i, $i = 1,...,m$, where m is the number of classes), each of which contains vertices (v_i) with equal distance values ($d(v_i)$). As a result, we obtain a spectrum of the distance parameters of the molecular graph in which distance values are plotted on the abscissa, and the number of vertices with a given distance are plotted on the ordinate.[12,13] The probabilities of the appearance of vertices of each kind can be calculated from this spectrum, as in the case of the mass spectrum. Substituting the values of these probabilities in the Shannon formula, we obtain the information topological index of the molecular graph H_{MG}, which is a quantitative measure of the complexity of the molecular structure and is expressed in the information units, bits (Fig. 1).

Mass spectral studies show that a richer fragmentation scheme corresponds, as a rule, to a molecule with a more complicated structure. Therefore a correlation between the information indices of the structure and the mass spectrum would be expected. In fact, these correlations were found in representative selections of about 20 classes of organic and organometallic compounds.[13—15] In particular, regression analysis of structural and mass spectral indices of arylsilanes (Fig. 2) revealed three groups of compounds, each of which is characterized by its structural peculiarities. Phenylmethylsilanes are in the first group, the second group contains allylphenyl- and phenylvinylsilanes, and methylvinylsilanes are in the third group.[14] The linear dependence of the information topological indices of molecular graphs of monosubstituted ferrocenes on the information indices of the mass spectra of these compounds is presented in Fig. 3. The whole set of compounds studied is divided into two subsets, each of which is characterized by its regression line with approximately the same slope.[13]

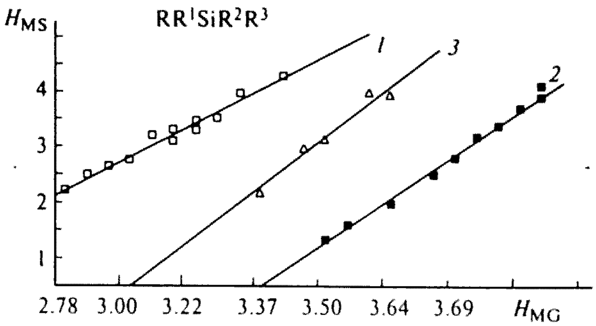


Fig. 2. Interrelation between mass spectra (H_{MS}) and structural (H_{MG}) indices of arylsilanes:

1, $H_{MS} = -5.9 + 2.9H_{MG}$ ($r = 0.945$, $n = 12$, R^1 = Ph, R^2 = Me);

2, $H_{MS} = -8.3 + 2.8H_{MG}$ ($r = 0.982$, $n = 10$, R^1 = All, Vin, R^2 = Ph);

3, $H_{MS} = -8.5 + 3.4H_{MG}$ ($r = 0.856$, $n = 6$, R^1 = Vin, R^2 = Me).

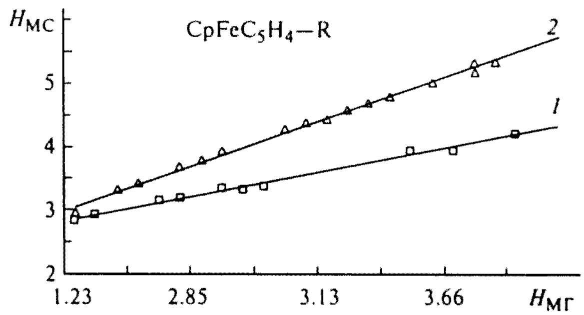


Fig. 3. Interrelation between structural (H_{MG}) and mass spectral (H_{MS}) indices of monosubstituted ferrocenes:

1, $H_{MS} = 1.70 + 0.57H_{MG}$ ($r = 0.940$, $n = 10$);

2, $H_{MS} = 1.23 + 0.98H_{MG}$ ($r = 0.975$, $n = 16$).

The existence of several dependences within one class of compounds is related to the fact that, in terms of the model used, information topological indices do not take into account the electronic effects of substituents, which can exert a substantial effect on the degree of fragmentation, as has been discussed above. This is also indicated by the linear correlations between the information indices of mass spectra and Hammett σ constants, which have been established for 10 classes of aromatic compounds (azo compounds, Schiff's bases, amides, and others).[15]

The results obtained indicate that the approach suggested is promising for the solution of the structure—reactivity problem. The correlations of information indices of mass spectra with Hammett σ constants are of special interest, because they allow dimensionless values of σ constants to be expressed in information units.

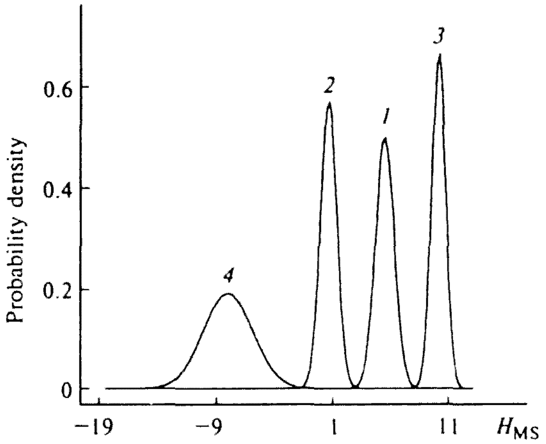


Fig. 4. Distribution of mass spectral indices (H_{MS}) for four classes of organometallic compounds:
1, $CpFeC_5H_4R$; *2*, $RC_5H_4Mn(CO)_3$; *3*, $RC_5H_4Re(CO)_3$; *4*, $RC_6H_5Cr(CO)_3$.

This, in turn, makes it possible to compare the extents of the effects of structural topological and electronic parameters of compounds on the processes of their fragmentation under electron impact in a common dimension scale.

Along with the estimation of the reactivity of various classes of compounds in a common dimension scale, the dependences established between mass spectral and structural indices can also be used to identify substances. Therefore, we developed a general approach to the formation of a system of mass spectral indices that unambiguously characterize the compound studied.[8] For example, the distribution of one of the mass spectral superindices H_{MS} is presented in Fig. 4 for four structurally related classes of organometallic compounds of the ferrocene (*1*), cymanthrene (*2*), cyclopentadienylrheniumtricarbonyl (*3*), and benzenechromiumtricarbonyl (*4*) series. It can be seen that each of them is characterized by the range of variation of its index that does not overlap other ranges, which makes it possible to use it for the unambiguous assignment of an unknown substance to one of these classes. Using a representative selection of one thousand five hundred organometallic compounds, it has been shown, in particular, that two mass spectral indices are enough for the unambiguous identification of an unknown compound whose mass spectrum is lacking in the library of spectra.[8]

Thus, the information approach suggested for interpretation of mass spectra allows one, on the one hand, to study the relation between the structure and reactivity of ions in the gas phase and, on the other hand, to identify unknown compounds.

In conclusion, let us briefly formulate the main statements of this approach: 1) the gas-phase fragmentation of molecules under electron impact is suggested as a standard reaction series for the quantitative estimation of the reactivity of all volatiles; 2) information indices of mass spectra, which take into account all possible directions of fragmentation of a compound under electron impact and which are a measure of the complexity of its intramolecular decomposition scheme, can serve as quantitative parameters of the reactivity of molecules in the gas phase; 3) information topological indices of the corresponding molecular graphs, which are a measure of the structural complexity of molecules, can serve as a generalized structural parameter of molecules; and 4) the structure (information topological indices of molecular graphs) and the reactivity (information indices of mass spectra) are compared in a common dimension scale (in the information units, bits).

This work was financially supported by the Russian Foundation for Basic Research (Projects Nos. 94-03-08126 and 96-03-33392).

References

1. J. I. Brauman, C. A. Leider, and W. N. Olmstead, *J. Am. Chem. Soc.*, 1974, **96**, 4030.
2. I. A. Koppel', U. Kh. Mel'der, and R. I. Pikver, *Elektronnoe i protonnoe srodstvo molekul* [*Electron and Proton Affinity of Molecules*], BNTs UrO AN SSSR, Ufa, 1991, 3 (in Russian).
3. Yu. S. Nekrasov and D. V. Zagorevskii, *Org. Mass Spectrom.*, 1991, **26**, 733.
4. M. M. Bursey, *Org. Mass Spectrom.*, 1968, **1**, 31.
5. Yu. S. Nekrasov and N. I. Vasyukova, *Org. Mass Spectrom.*, 1979, **14**, 422.
6. Yu. S. Nekrasov, D. V. Zagorevskii, and M. Sh. Mikeladze, *Metalloorg. Khimi.*, 1992, **5**, 499 [*Organomet. Chem. USSR*, 1992, **5** (Engl. Transl.)].
7. K. Levsen, *Fundamental Aspects of Organic Mass Spectrometry*, Weinheim, Verlag Chemie, 1978, 152.
8. Yu. N. Sukharev, Yu. S. Nekrasov, N. S. Molgachova, and E. E. Tepfer, *Org. Mass Spectrom.*, 1993, **28**, 1555.
9. D. Bonchev, *Information Theoretic Indices for Characterization of Chemical Structures*, New York: Research Studies Press, 1983, 249 pp.
10. D. H. Rouvray, in *Chemical Applications of Topology and Graph Theory*, Ed. R. B. King, Elsevier, Amsterdam , 1983].
11. M. I. Stankevich, I. V. Stankevich, and N. S. Zefirov, *Usp. Khim.*, 1988, **57**, 337 [*Russ. Chem. Rev.*, 1988, **57**, 191 (Engl. Transl.)].
12. Yu. S. Nekrasov, E. E. Tepfer, Yu. S. Sukharev, V. A. Skorobogatov, and E. V. Konstantinova, in *Teoriya grafov i ee prilozheniya* [*Graph Theory and its Applications*], Novosibirsk, 1994, no. 151: *Vychislitel'nye sistemy* [*Calculation Systems*], 88 (in Russian).
13. V. A. Skorobogatov, E. V. Konstantinova, Yu. S. Nekrasov, Yu. N. Sukharev, and E. E. Tepfer, *MATCH*, 1991, **26**, 215.
14. Yu. S. Nekrasov, E. E. Tepfer, and Yu. N. Sukharev, *Izv. Akad. Nauk, Ser. Khim.*, 1993, 381 [*Russ. Chem. Bull.*, 1993, **42**, 343 (Engl. Transl.)].
15. Yu. S. Nekrasov, E. E. Tepfer, and Yu. N. Sukharev, *Izv. Akad. Nauk, Ser. Khim.*, 1993, 2073 [*Russ. Chem. Bull.*, 1993, **42**, 1986 (Engl. Transl.)].

Received June 27, 1996