Russian Journal of Organic Chemistry, Vol. 40, No. 5, 2004, pp. 644–649. Translated from Zhurnal Organicheskoi Khimii, Vol. 40, No. 5, 2004, pp. 675–680.

Fragment Descriptors in QSPR: Application to Heat Capacity Calculation

A.A. Ivanova, V.A. Palyulin, A.N. Zefirov, and N.S. Zefirov

Moscow State University, 119992, Moscow, Russia
e-mail: ivanova@org.chem.msu.su; zefirov@org.chem.msu.ru

Received December 20, 2003

Abstract—Within the framework of QSPR simulation methodology the practicality was demonstrated of fragment approach for forecasting heat capacity of various class organic compounds. An equation was advanced permitting reliable prediction of organic compounds heat capacity proceeding from V_x, parameter characterizing the molecule volume, and from fragment descriptors.

Structural theory is the foundation of organic chemistry. "It is the basis for systematization of million facts about hundred thousands of individual compounds" [1]. The core of the structural theory is the structural formula: just the establishment of the structural formula of a chemical compounds provides the knowledge on its chemical behavior. Within the framework of the structural theory many important concepts have been developed, among which we wish to point out ***homology*** and ***functional group***. The latter is defined as a part of a structural formula that bears a definite set of qualities, and moreover, this set possesses transferability. These concepts were the first basis for rationalizing the dependence of the properties of organic compounds both on molecular weight and on the presence in the molecule of atomic sets bound in a special way. Actually, within this framework the properties of organic substances may be regarded as dependent on the molecular "skeleton" and on the presence of certain functional groups (or "substituents"), every one of which contributes to the overall properties.

Moreover, the application of additivity schemes in the study of properties was quite natural in the framework of the structural theory. In this case a general property is described as a sum of contributions from individual atoms or bonds or groups of atoms bound together (structural fragments) by an equation

$$A = \sum_{1}^{N} A_i \cdot n_i \qquad (1)$$

where A is a value of a property of compound, n_i is a number of structural fragments of a certain kind (depending on the given scheme), A_i is the contribution of the corresponding fragment into the property, and N is the total number of fragment types.

We can mention as examples the classic additive schemes for calculating molecular refraction {both through atoms [2, 3] and bonds [4–6]), parachor [7, 8], heats of combustion and formation [9–13], lipophily [14–17] etc. The additive schemes were often unsatisfactory (limited to certain classes of compounds, possessed poor prediction possibility etc.). As a result they were growing more complicated; probably the most intricate systematism was developed in the works of Tatevskii and his team [18].

Actually, the same reasoning underlies the Hammett's equation: a general property is regarded as a characteristic of a certain "core" and it is modified by "substituents", and to each substituent is inherent some partial value of the property. These concepts were developed by Hansch [14, 15], that further by invoking the approaches based on graph theory resulted in development of QSAR [19].

It is possible that Smolenskii has been the first to use the subgraph sets for calculation of physicochemical properties [20, 21]. A property was regarded as a linear function of many variables, every one among which was a number of specific subgraphs (chains) in the molecular graph. By and large, all fragment methods are based on the general formula:

$$A = A_0 + \sum_{1}^{N} A_i \cdot n_i \qquad (2)$$

1070-4280/04/4005-0644©2004 MAIK "Nauka/Interperiodica"

where the overall property A is represented as a sum of properties of fragments (that may be overlapping) added to a certain constant A_0. However the main development of QSAR was connected with the use of topologic and quantum-chemical descriptors of molecular structure, and also employing physicochemical characteristics (refraction, lipophily etc.).

The attempts to use in QSAR fragments (subgraphs) for description of molecular structure (formula) originate from the general logic of the structural theory. This concept finds support also in the graph theory: it was shown that **any topologic index may be replaced by a set of substructural descriptors if the database "structure-property" is sufficiently large for building up a statistically valid model** [22].

Yet the existing fragment approaches differ in their definition of structural fragments and ways and strategy of their generation. The most well-known approach is the Free-Wilson method underlain by linear regression analysis using as descrptors "indicator variables" (showing the presence on lack of a definite fragment) [23]. As fragments serve usually simple substituents containing a small number of atoms of different types [24]. The application of this methodology (even in a very simplified form) together with topologic indices from the graph theory or with some experimental parameters (e.g., *logP*) also proved to be sufficiently successful.

On the other hand, the fragment approach was transformed into "fingerprints" analysis [25]. Frequently for this purpose are used relatively simple, often unmarked subgraphs. For QSAR purposes all these descriptors are regarded as variables for various statistical treatments, like linear regression (LR), multiple linear regression (MLR), partial least-squares method (PLS), artificial neuron nets (ANN), etc. The fragment approach of Klopman (CASE) [26] and Meylan-Howard (AFC, atomic fragments contribution) [27] should be separately considered. Both approaches approximate a molecular property by summation of local contributions of various fragments, whereas the fragment contribution depends on its local neighbor in the molecule. In Klopman approximation mainly complex fragments are used that sometimes are associated with biophores and biophobe sconcepts. In the Meylan-Howard approach monatomic fragments are mainly applied, whereas the complex nature of the molecular structure is accounted for by a number of correcting factors. Estrada publications [28, 29] also are worth mentioning. Here the spectral moments of lines contact matrix in the molecular graph are represented as linear combination of various structural fragments. These descriptors provided a possibility to build up QSPR models for a number of physicochemical properties of individual classes of organic compounds.

In 1990 we developed a subroutine FRAGMENT capable of generating sets of fragments: chains (1–9 atoms), rings (3–6-membered), and several types of branched fragments [30–35]. Moreover, each atom in the fragment was coded depending on its type, surrounding bonds, and number of neighboring hydrogens thus providing a possibility to flexibly account for multiple bonds, functions, heteroatoms etc. The subroutine FRAGMENT was successfully used in the program QSAR/QSPR complex EMMA [30–35], and also in the neuron net complex NASAWIN [35, 36]. We recently completed the development of a refined version of subroutine FRAGMENT distinguished by a large number of fragment types and by more flexible classification of atoms [37].

Our previous studies in the QSAR/QSPR field tested the applicability of fragment descriptors (either *per se* or combined with other descriptors) for building up QSAR/QSPR models of physicochemical and biological properties of organic compounds [38–42]. The advantage of fragment descriptors lies in a pictorial rendition and easy structural interpretation of QSAR/QSPR results. The present study was aimed at investigation of applicability of structural descriptors to QSPR processing of organic compounds heat capacity.

The heat capacity is defined as a quantity of heat required for raising the temperature by 1°C. The heat capacity data are necessary for calculation of thermodynamic functions, and in practice the heat capacity evaluation is important for calculation of energy balance of chemical processes in reactors, for selection of optimum heat transfer agents, etc. The experimental measurement of heat capacity is relatively labor-consuming and expensive procedure, and therefore appear QSAR/QSPR data on simulation of this property [43–45]. Gakh applied the method of artificial neuron nets to prediction of alkanes heat capacity (*RMS* 4.04) [44]. An extensive and versatile database on heat capacity for various classes of organic compounds is presented in [45] , and a good QSPR model is constructed using neuron nets (for complete sampling *RMS* 17.141, for teaching sampling *RMS* 16.857, and for control sampling *RMS* 18.744).

EXPERIMENTAL

The database (DB) for the present study was taken from [45]. It includes 871 heat capacity values for versatile classes of organic compounds, like alkanes,

alkenes, alkynes, dienes, cycloalkanes, cycloalkenes, haloalkanes, haloalkenes, alcohols, carboxylic acids, carbonyl compounds, esters, amines, aromatic amines, nitriles, pyridines, sulfides, thiophenes, polyfunctional and various aromatic compounds. Hence we deal with a structurally dissimilar access. The total volume of DB is 871 compound. For building up QSPR models of DB we developed special programs MEOW and BASTET.*

These programs were used for checking the BD compiled in [45], and errors and double entering of structures was revealed. The personal contacts with Xiaojun et al. permitted to make the following corrections in the DB: 1. Compound numbered 235 should be 3-methyl-1,2-butadiene, heat capacity 152.08. 2. Compound numbered 528 should be octane, heat capacity 254.71. 3. Compound numbered 515 should be 1-diacetoxy-2-ethoxyethane, heat capacity 300.32. 4. Compound numbered 821, described as nonylphenol is the *ortho*-isomer. We used in the study program QSAR/QSPR complex EMMA developed at the Chemical Department of the Moscow State University [30–35], neuron net NASAWIN [35, 36], and combined therewith refined version of routine FRAGMENT.

In the first stage of the study we performed a search for uniparametric correlation with the use of all descriptors involved in EMMA program. The most statistically significant proved to be descriptor V_x [46, 47] that was the measure of molecular volume. This descriptor was calculated according to the formula.

$$V_x = 16.35N_C + 8.71N_H + 12.43N_O + 14.39N_N + 10.48N_F + 20.95N_{Cl} + 6.21N_{Br} + 34.53N_I + 22.91N_S + 24.87N_P - 6.56N_b \quad (3)$$

where N_C, N_H, N_O, N_N, N_F, N_{Cl}, N_{Br}, N_I, N_S, N_P are numbers of atoms of the corresponding element contained in the molecule, and N_b is the number of bonds in the molecule (it should be noted that the above formula is valid only for compounds containing the listed types of atoms).

The uniparametric QSPR equation $C = 3.3V_x + 6.86$ obtained has the following statistical parameters for the teaching sampling R^2 0.9359, s 27.4, F 10858, maximum error (for oleic acid) 121.6 J mol^{-1} K^{-1}, and for the control sampling R^2 0.9381, average error is equal to 20.8, maximum error (for 1,3-propylene glycol) 103.7 J mol^{-1} K^{-1}.

In the next stage of the study intermediate models were constructed in order to facilitate the choice of descriptors for simulating the total DB. To this end we built up models basing on V_x index with stepwise addition of fragment descriptors (Table 1).

Further individual models were constructed for the following groups of compounds: hydrocarbons, nitrogen-containing, sulfur-containing, oxygen-containing, halogen-containing, and bifunctional compounds (Table 2).

Taking into consideration all the data obtained we built up a model "structure–property" for total DB. In constructing the general model we used V_x index and the set of fragment descriptors found in modeling of heat capacity for compounds groups. The constructed QSPR model (equation 4) possesses the following statistical parameters: R^2 0.9787, s 15.86, F 5667 (Fig. 1), and for the control sampling the average error is equal to 13.081 J mol^{-1} K^{-1}, R^2 0.9750, F 769 (Fig. 2). In the QSPR equation obtained is included V_x index and five fragment descriptors characteristic of the compounds classes involved in the access:

$$C = -3.942 + 1.944V_x + 27.681D2 + 11.655D3 + 3.778D4 + 26.985\,D5 + 10.864\,D18 \quad (4)$$

As follows from the values of Student criterion (Table 3) the contribution of V_x index into the QSPR model obtained is the largest. Into the final QSPR equation is beside included a set of fragment descriptors for structural features of individual compounds classes. For instance, in presentation of oxygen-containing compounds descriptors $D2$ and, $D3$ are employed taking into account respectively the number of oxygens atoms and hydroxy groups in the molecule. The application of two descriptors for the oxygen-containing compounds is due to versatility of structures of this type occurring in DB (alcohols, carboxylic acids, esters, aldehydes and ketones, and nitrocompounds). The signigicance of these descriptors also is confirmed by the high values of Student criterion (Table 3).

The next in significance descriptor according to the obtained Student criterion value is $D18$ accounting for the number of fluorine atoms in a molecule. Apparently this descriptor gains its significance from the fact, that the fluorine-containing compounds amount to 4% from the total DB and 29% from the overall number of halogen derivatives. On exclusion of this descriptor from the obtained QSPR equation the R^2 decreases to 0.9700, and s increases to 17.70.

Descriptor $D4$ accounting for the number of of >CH_2 groups in a molecule also plays a significant part in the

* Authors can supply programs MEOW and BASTET on special request.

Table 1. Statistical parameters of QSPR model for heat capacity based on V_x index and fragment descriptors

Number of descriptors	1	2	3	4	5	6	7
Descriptors of model	*D*1	*D*1, *D*2	*D*1, *D*2, *D*3	*D*1, *D*2, *D*3, *D*4	*D*1, *D*2, *D*3, *D*4, *D*5	*D*1, *D*2, *D*3, *D*4, *D*5, *D*6	*D*1, *D*2, *D*3, *D*4, *D*5, *D*6, *D*7
Teaching sampling 746 compounds							
R^2	0.9359	0.9585	0.9643	0.9702	0.9735	0.9754	0.9757
s	27.4	22.1	20.5	18.7	17.7	17.0	16.6
F	10858	8589	6676	6035	5431	4884	4412
Maximum error	121.6(734)	88.8(74)	93.0(74)	95.8(125)	97.2(125)	76.5(635)	76.6(635)
Control sampling 125 compounds							
R^2	0.9381	0.9573	0.9647	0.9679	0.9700	0.9678	0.9702
Average error on a prediction	20.7	17.2	15.2	14.1	12.5	14.3	13.4

Note: Descriptor *D*1 is index V_x, *D*2 is number of OH groups, *D*3 is number of oxygen atoms, *D*4 is number of $>CH_2$ groups, *D*5 is number of –N fragments, *D*6 is number of fragments ⟩–⟨, *D*7 is number of fragments CH_2CH_2CH; 734 is oleic acid, 74 is octafluoropropane, 125 is decafluorobutane, **635** is *n*-butyl benzoate.

Table 2. Statistical parameters of QSPR model for heat capacity based on V_x index and fragment descriptors

Group of compounds	Equation and statistical parameters of QSPR models
Hydrocarbons	$C = -7.293 + 2.021D1 + 2.456D4 - 4.161D8$, $N = 298$, $R^2 = 0.9866$, $s = 11.0$, $F = 7210$
Sulfur-containing compounds	$C = -7.846 + 2.201D1 - 4.108D9 - 35.313D10$, $N = 79$, $R^2 = 0.9943$, $s = 11.0$, $F = 4392$
Halogen-containing compounds	$C = 20.067 + 1.581D1 + 12.107D18 - 4.487D19 - 15.087D20$, $N = 95$, $R^2 = 0.8866$, $s = 13.2$, $F = 176$
Nitrogen-containing compounds	$C = 13.805 + 1.949D1 + 16.500D11 + 6.437D12$, $N = 51$, $R^2 = 0.9660$, $s = 12.7$, $F = 445$
Oxygen-containing compounds	$C = -12.132 + 2.332D1 - 3.861D13 + 4.943D14 + 32.847D2$, $N = 189$, $R^2 = 0.9751$, $s = 16.3$, $F = 1801$
Bifunctional compounds	$C = 53.271 + 1.317D1 - 22.189D15 + 29.302D16 + 25.230D17$, $N = 34$, $R^2 = 0.9015$, $s = 18.0$, $F = 66$

Note. Descriptor *D*8 is number of fragments CH_2–CH, *D*9 is number of fragments C=, *D*10 is number of fragments S–S, *D*11 is number of NH_2 groups, *D*12 is number of fragments $–C_{sp^3}–N_{sp^3}$, *D*13 is number of fragments C_{Ar}, *D*14 is number of fragments =CH, *D*15 is number of $–CH_3$ groups, *D*16 is number of fragments C–O, *D*17 is number of fragments C–N, *D*18 is number of fluorine atoms, *D*19 is number of fragments CC(C)H, *D*20 is number of fragments $C_{sp^3}Cl_2F$.

obtained QSPR model. The role of this descriptor in QSPR equation is multipurpose: it suits for making an allowance for the length of unsubstituted saturated hydrocarbon chain, has different values for alkanes and alkenes with the same number of carbon atoms, and often has different value for isomers (for instance, for 1-butene *D*4 = 1, and for 2-butene *D*4 = 0; for 1-propanol descriptor *D*4 is 2, whereas for 2-propanol it takes a zero value).

Table 3. Student criterion values for descriptors of QSPR model

	Descriptor	Student criterion
V_x	V_x	85.472
*D*3	O	16.849
*D*2	–OH	16.153
*D*18	F	13.513
*D*4	CH_2	12.983
*D*5	–N	11.589

Descriptor *D*5 included in the equation obtained and accounting for the number of nitrogen atoms also corresponds to sufficiently high value of Student criterion (*t* 11.6). It is worth mentioning that the nitrogen-containing compounds amount to 9% of the total number of compounds included in DB. It is apparently just the reason why this descriptor significantly contributes to the regression equation obtained. It is also necessary for description of polyfunctional nitrogen-containing compounds.

Alongside the linear regression analysis we also performed simulation with the use of artificial neuron nets (ANN). The neuron net model was constructed using

the descriptors involved in equation (4). The input layer of the net was composed of seven neurons in keeping with the number of the preliminary selected descriptors, the hidden layer contained three sigmoid neurons, and the output layer consisted of a single neuron corresponding to the predicted property. As teaching algorithm was taken the "generalized delta-rule" [48], parameter of teaching velocity was 0.25, the value of teaching "moment" parameter was 0.9. The teaching process was stopped when the least prediction error for the control sampling was attained. The obtained QSPR model has the following statistical parameters: R^2 0.9795, *RMS* for teaching sampling 15.7, *RMS* for control sampling 18.5. The comparison of result obtained with that of [45] shows some refinement of *RMS* value probably because of removing errors and double structures from DB. On the other hand, the linear regression model we obtained is distinguished by sufficiently large correlation factor.

Thus we demonstrated that the heat capacity of the most classes of organic compounds well correlated with V_x index and fragment descriptors. The fragment descriptors are easy to calculate and to interpret, and therefore they can be used in QSPR analysis. It should be noted that the general model for heat capacity calculation of organic compounds developed in the present study is based on taking into account small fragments and therefore is sufficiently versatile and applicable to molecules of various dimensions.

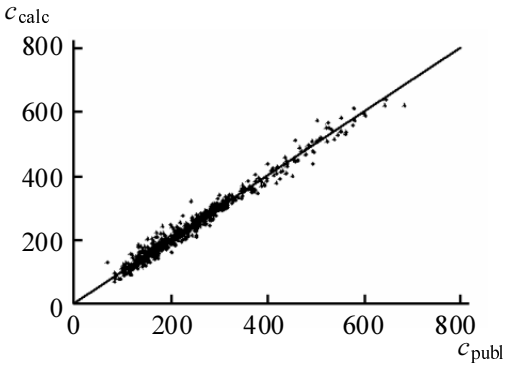

Fig. 1. QSPR model for heat capacity of various classes of organic compounds.

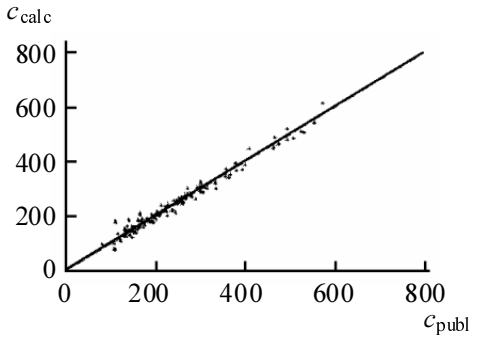

Fig 2. Control sampling for the QSPR model obtained.

REFERENCES

1. Morrison, R.T. and Boyd, R.N., *Organic Chemistry,* Boston: Allyn and Bacon Inc., 1983. Ch. 1, 2.
2. Reinhard, M. and Drefahl. A., *Handbook for Estimating Physicochemical Properties of Organic Compounds*, New York: Wiley, 1999.
3. Wildman, S.A. and Grippen, G.M., *J. Chem. Inf. Comput. Sci*., 1999, vol. 39, p. 868.
4. Cox, J.D. and Pilcher, G., *Thermochemistry of Organic and Organometallic Compounds,* New York: Academic, Press, 1970.
5. Vogel, A.I., Cresswell, W.T., Jeffery, G.H., and Leicester, J., *J. Chem. Soc*., 1952, p. 514.
6. Batsanov, S.S., *Strukturnaya, refraktometriya* (Structural Refractometry), Moscow: Izd. Mosk. Gos. Univ., 1976.
7. Gibling, T.W., *J. Chem. Soc.,* 1943, p. 146.
8. Samuel, R., *J. Chem. Phys*., 1944, vol. 12, p. 167.
9. Benson, S.W. and Buss, J.H., *J. Chem. Phys*., 1958, vol. 29, p. 546.
10. Bernstein, H.J., *J. Chem. Phys*., 1952, vol. 20, p. 263.
11. Franklin, J.L,. *J. Chem. Phys*., 1952, vol. 21, p. 2029.
12. Sounders, M., Matthews, C.S., and Hurd, C.O., *Ind. Eng. Chem*., 1949, vol. 41, p. 1037.
13. Zahn, C.T., *J. Chem. Phys*., 1934, vol. 2, p. 671.
14. Fujita, T., Iwasa, J., and Hansch, C., *J. Am. Chem. Soc*., 1964, vol. 86, p. 5175.
15. Hansch, C., Muir, R.M., Fujit, T., Maloney, P.P., Streich, M., *J. Am. Chem. Soc*., 1963, vol. 85, p. 2817.
16. Nys, G.G. and Rekker, R.F., *Eur. J. Med. Chem. Chim. Therap.,* 1973, vol. 8, p. 521.
17. Rekker, R.F. and de Kort, H.M., *Eur. J. Med. Chem*., 1979, vol. 14, p. 479.
18. Tatevskii, V.M., *Teoriya fiziko-khimicheskikh svoistv molekul i veshchestv* (Theory Phisicochemical Properties of Molecules and Substances), Moscow: Izd. Mosk. Gos. Univ., 1987; Papulov, Yu.G., *Molekulyarnye grafy* (Molecular Graph), Tver': Izd. Tver. Gos. Univ., 1990.
19. Kubinyi, H., *Burger's Medicinal Chemistry and Drug Discovery,* Wolff, M.E., Ed., New York: Wiley, 1995, vol. 1, p. 497.
20. Smolensky, E.A., *Zh. Fiz. Khim*., 1964, vol. 38, p. 1288.
21. Smolenskii, E.A., *Dokl. Akad. Nauk, SSSR*., 1976, 230, p. 373.
22. Baskin, I.I., Skvortsova, M.I., Stankevich, I.V., and Zefirov, N.S., *J. Chem. Inf. Comput. Sci*., 1995, vol. 35, p. 527.
23. Free, S.M., Wilson, J.W., *J. Med. Chem*., 1964, vol. 7, p. 395.
24. Kubinyi, H., *QSAR: Hansch Analysis and Related*

Approaches, Mannhold, R., Krogsgaard-Larsen, P., and Timmerman, H.,. Eds., Germany, Weinheim: VCH, 1993.
25. Butina, D.J., *J. Chem. Comput. Sci., 199*9, vol. 39, p. 747.
26. Klopman, G., *J. Am. Chem. Sci*., 1984, vol. 106, p. 7315.
27. Meylan, W.M. and Howard, P.H., *J. Pharm. Sci*., 1995, vol. 84, p. 83.
28. Estrada, E., *J. Chem. Comput. Sci*., 1996, vol. 36, p. 844.
29. Estrada, E., *J. Chem. Comput. Sci*., 1997, vol. 37, p. 320.
30. Baskin, I.I., Palyulin, V.A., and Zefirov, N.S., Abstracts of Papers, Konf. "Molekulyarnye grafy v khimicheskikh issledovaniyakh" (Moleculaler Graphs in Chemical Researches), Kalinin, 1990, p. 5.
31. Palyulin, V.A., Baskin, I.I., Petelin, D., and Zefirov, N.S., Abstracts of Papers, *10th Europ Symp Structure-Activity Relationships: QSAR and Molecular Modelling*, Barcelona, 1994, B257.
32. Palyulin, V.A., Baskin, I.I., Petelin, D.E., and Zefirov, N.S., *QSAR and Molecular Modelling: Concepts Computational Tools and Biological Applications*, Sanz, F., Giraldo, J., and Manaut, F., Barcelona: Prous Science Publishers, 1995, p. 51.
33. Zefirov, N.S., Petelin, V.A., Palyulin, V.A., and McFarland, J., *Dokl. Akad. Nauk SSSR.,* 1992, vol. 327, p. 504.
34. Zefirov, N.S. and Palyulin, V.A., *J. Chem. Inf. Comput. Sci*., 2002, vol. 41, p. 1112.
35. Baskin, I.I., Palyulin, V.A., and Zefirov, N.S., *J. Chem. Inf. Comput. Sci*., 1997, vol. 37, p. 715.
36. Baskin, I.I., Palyulin, V.A., and Zefirov, N.S., Abstracts of Papers, *10th Europ. Symp. Structure-Activity Relationships: QSAR and Molecular Modelling,* Barcelona, 1995, p. 30.
37. Artemenko, N.V., Baskin, I.I., Palyulin, V.A., and Zefirov, N.S., *Dokl. Akad. Nauk.,* 2001, vol. 381, p. 203.
38. Pivina, T.S., Sukhachev, D.V., and Maslova, L.K., *Dokl. Akad. Nauk SSSR*., 1993, vol. 330, p. 468.
39. Sukhachev, D.V., Pivina, T.S., Zhokhova, N.I., Zefirov, N.S., and Zeman, S.I., *Izv. Akad. Nauk, Ser. Khim*., 1995, p. 1653.
40. Sukhachev, D.V., Pivina, T.S., Zhokhova, N.I., and Zefirov, N.S., *Izv. Akad. Nauk, Ser Khim*., 1995, p. 1657.
41. Zefirov, N.S., Palyulin, V.A., Oliferenko, A.A., Ivanova, A.A., and Ivanov, A.A., *Dokl. Akad. Nauk*, 2001, vol. 381, p. 637.
42. Zefirov, N.S. and Palyulin, V. J., *Chem. Inf. Comput. Sci*., 2001, vol. 41, p. 1022.
43. Yaws, C.L., *Chemical Properties Handbook*, New York: McGraw-Hill, 1999.
44. Gakh, A.A., Gakh, E.G., Sumpter, B.G., and Noid, D.W., *J. Chem. Inf. Comput. Sci*., 1994, vol. 34, p. 832.
45. Xiaojun, Yao, Botao, Fan, Doucet, J.P., Panaye, A., Mancang, Liu, Ruisheng, Zhang, Xiaoyun, Zhang, and Zhide, Hu, *QSAR Comb. Sci.,* 2003, vol. 22, p. 29.
46. Van de Waterbeemd, H. and Testa, B., *Adv. Drug Res*., Testa B, New York: Academic, Press, 1987, vol. 16, p. 85.
47. Abraham, M.H. and McGowan, J.C. *Chromatographia*, 1987, vol. 23, p. 243.
48. Zupan, J. and Gasteiger, J., *Neural Networks for Chemists – An Introduction*, Germany, Weinheim: VCH, 1993.