REVIEWS

Methodological Approaches to the Calculation and Prediction of Retention Indices in Capillary Gas Chromatography

A. K. Zhokhov[a], A. Yu. Loskutov[a], and I. V. Rybal'chenko[b], *

*[a]33rd Central Research and Testing Institute, Ministry of Defense of Russian Federation,
Volsk-18, Saratov oblast, 412918 Russia*

[b]27th Scientific Center, Ministry of Defense of Russian Federation, Moscow, 105005 Russia

**e-mail: rivrus@mail.ru*

Received January 24, 2017; in final form, August 27, 2017

Abstract—The scientific achievements in the field of calculation and prediction of gas chromatographic indices of retention of organic compounds on standard nonpolar polydimethylsiloxane stationary phases of capillary chromatographic columns are analyzed. Approaches based on the study of correlation dependences with the use of physicochemical, fragment, and quantum-chemical descriptors, topological indices, and various additive models are described.

Gas chromatography using standard high-performance capillary columns is recognized as an effective method for qualitative and quantitative determination of a wide range of organic compounds [1]. A good reproducibility of the retention parameters in capillary gas chromatography arouse interest of researchers in using the retention indices of compounds to be determined as unique characteristics for identifying the components of mixtures being analyzed [2].

According to the definition of Kovats [3, 4], the retention index (RI, designation in equations *Ri*) is a measure of the relative retention of matter, with a hydrocarbon of normal structure being used as the reference substances. A retention index is assigned to each normal hydrocarbon, which is the number of carbon atoms in its molecule multiplied by 100. For example, the retention indices for *n*-pentane and *n*-decane are 500 and 1000, respectively. Chapter 3.7.15 of *IUPAC Chromatographic Nomenclature* [5] defines that the Kovats' retention index is the result of logarithmic interpolation, obtained in measurements under isothermal conditions:

$$Ri = 100n + 100\left(\frac{\log t_a - \log t_n}{\log t_{(n+1)} - \log t_n}\right), \tag{1}$$

where *Ri* is the retention index; *n* is the number of carbon atoms in the *n*-alkane; $\log t_a$, $\log t_n$, and $\log t_{(n+1)}$ are the logarithms of the retention time of a component under study and two hydrocarbons, between which the component elutes.

When analyzing mixtures containing components with a wide range of boiling points, temperature programming is generally used. With linear temperature programming, the retention times of the components of a homologous series increase linearly with increasing number of carbon atoms. In the case of temperature programming, the corresponding quantities are called "linear retention indices."

Van den Dool and Kratz proposed a method of using retention indices in gas chromatography with temperature programming [6]. According to their approach, the calculation of retention indices is simplified since linear interpolation is performed, and the retention time values can be used, rather than their logarithms, that is,

$$Ri = 100n + 100\left(\frac{t_a - t_n}{t_{(n+1)} - t_n}\right). \tag{2}$$

Other compounds can be used as reference substances instead of *n*-alkanes. For example, in the identification of phosphorus-containing organoelement compounds, series of compounds with the following general structures (I, II, III) were proposed as reference substances for the determination of retention indices [7, 8]:
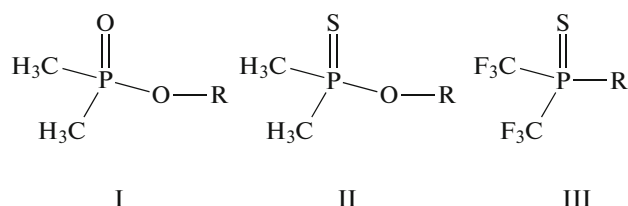
where R is an acyclic linear alkyl radical containing from 1 to 15 carbon atoms in the structure; the alkyl radical in compounds III contains from 3 to 20 carbon atoms.

The retention indices do not depend on the errors in measuring the flow rate of the mobile phase or the adsorbent weight in the column, and they are relatively insensitive to temperature changes. In using them, no correction for pressure drop is required [9, 10]. The use of retention indices avoids the errors of the first and second kind, which appear in the identification of isomers by their mass spectra, because of the insensitivity of these quantities to the isomerism of the carbon skeleton [11, 12].

In present-day automated systems for the identification of organic compounds by gas chromatography in combination detection by electron ionization mass spectra, the retention characteristics are considered as independent of mass spectral ones; they are subject to accounting in assessing the probability of correct identification. For example, in the widely used Automated Deconvolution and Identification System software product [13], even when the mass spectrum of the compound under study is completely identical with that of the mass spectral library, the identification is not recognized as reliable if the difference between the experimental and library values of the retention index exceeds a predetermined value (±20 index units). The condition of the working capacity of such systems is, obviously, the "saturation" of all mass spectral libraries connected to them with the reference values of the retention indices.

Currently, the chromatography–mass spectrometric databases contain hundreds of thousands of reference mass spectra and, unfortunately, a significantly smaller number of reference values of chromatographic retention indices [14]. For example, the NIST-14 database contains 385 872 reference values of retention indices for 82 868 substances [15]. At the same time, for other substances, the total number of which in this mass spectral library is more than 250 000, the reference retention indices are not given at all. The absence of an exhaustive array of retention indices in the most widely used mass spectral libraries (NIST, Wiley) significantly limits the possibility of reliable identification of compounds, which makes it urgent to create a computational tool for their replenishment.

The goal of the present paper is to generalize the methods for predicting gas-chromatographic retention indices.

Analysis of the publications allows a conditional division of the methods for predicting gas-chromatographic retention indices into two groups: (1) the methods based on the search for correlation dependences between the retention parameters and the structural or physicochemical properties of the substances and (2) the approaches built on the use of additive models based on the classical theory of the structure of molecules.

Correlation dependences of the retention parameters on structural or physicochemical properties of substances. The models that enable prediction of the quantitative characteristics and physical and physicochemical properties of organic compounds gained an English-language name Quantitative Structure Property Relationship (QSPR). The study of chromatographic retention is a part of the model defined by the general name Quantitative Structure Activity Relationship (QSAR) [16]. Information on the application of QSPR/QSAR methods in the period from 1996 to 2006 is reviewed in [16]. The data obtained from 2007 to 2011 are summarized in [17], where practical advices are given on how to avoid typical problems when using models.

Molecular descriptors are used in these models to describe the structures of chemical compounds. Molecular descriptor is a parameter that characterizes the structure of an organic compound, based on identifying specific features of this structure. In principle, a descriptor can be any number that can be calculated from the structural formula of a chemical compound, for example, molecular weight, the number of atoms of a certain type (hybridization), bonds or groups, molecular volume, partial charges at atoms, etc. [18].

In QSPR/QSAR practice, fragment descriptors, topological indices, physicochemical descriptors, quantum chemical descriptors, etc. are used to predict chromatographic characteristics. A rather complete set of molecular descriptors used in modern structure–property studies is described in [19], which can be considered an encyclopedia of molecular descriptors. Consider the possibility of using molecular descriptors to predict the retention indices.

Application of physicochemical descriptors. Physicochemical descriptors [18] are numerical characteristics obtained in the simulation of the physicochemical properties of chemical compounds or quantities having a clear physicochemical interpretation. The method for prediction of the retention indices is based on the use of very detailed and well-systematized data on the physicochemical properties of organic compounds. These include boiling point, volatility, dipole moment, Henry coefficient, molar refraction, saturated vapor pressure, and other properties [20–25].

Review [26] considers the energy characteristics and physical properties of adsorbates used in the correlation ratios of the retention parameters. As energy characteristics, the authors used the electronic polar-

izability, the ionization potential, the diamagnetic susceptibility, the electron density of molecules, and the resonant energy of the π-electron system. The physical properties of adsorbates, which have found application in correlation ratios, include heat and energy of formation, surface tension, density, viscosity, heat of crystallization, evaporation, and combustion, and molar volume. Such characteristics are determined for a small number of compounds; therefore, the possibility of predicting retention parameters based on them is extremely limited.

The correlation three-parameter equations are characterized by the highest accuracy of the calculation of retention indices; they connect the chromatographic constants with the boiling points of organic substances [22]:

$$\log Ri = a \log T_{bp} + bA + c, \qquad (3)$$

where T_{bp} is the boiling points of organic substances of one homologous series; a, b, and c are parameters of the linear—logarithmic three-parameter equation; and A is a parameter corresponding to the taxonomic groups of homologues or congeners. In the case where the retention indices are calculated within the isomer groups, that is, parameter A has a constant value, we can use equation [27]

$$Ri = a T_{bp} + b, \qquad (4)$$

where T_{bp} is the boiling points of organic substances of one homologous series; and a and b are the parameters of the equation.

The number of parameters in the equations used is no less important. To calculate the retention indices by physicochemical constants, nonlinear four-parameter relations are proposed [25] in addition to the three-parameter equations (3).

The retention indices obtained by using chromatographic phases of different polarity can serve as physicochemical characteristics for the correlation of retention parameters [28]. Such methods of calculation are more accurate than the correlations using physicochemical properties, but at the same time, they are more limited in use in connection with the lack of experimental data for different chromatographic phases.

The disadvantage of the method of approximation of retention indices with the use of various physical and chemical quantities is the lack of a sufficient array of reference data relating to the compounds under study.

It is shown [22] that the physicochemical characteristics can be calculated using the principle of structural analogy. The retention indices in this case can be calculated from the boiling points of any structural analogs, that is,

$$\log Ri = a' \log T_{bp}^{*} + b'A + c', \qquad (5)$$

where T_{bp} is the boiling points of the structural analogue; a', b', and c' are parameters of the linear—logarithmic three-parameter equation; and A is a parameter corresponding to the taxonomic groups of homologues or congeners. Using the method of structural analogy, based on the study of regularities of the change in the properties of substances in the transition from one molecular structure to another, enables the identification of the patterns of changes in the physicochemical properties of compounds with a gradual complication of the structure of molecules and thus the determination of the characteristics of substances for that experimental data cannot be obtained for a number of reasons [22, 29].

The accuracy of calculations depends largely on the proper selection of structural analogues in each case under consideration [27]. As structural analogues, it is proposed to consider substances with a structurally close carbon skeleton and different functional groups, multirow homologs, and aromatic and heterocyclic compounds, that is, rather broad taxonomic groups. Either groups of substances that include a large number of diverse structures or, on the contrary, taxonomic groups containing a small number of relatively structurally similar substances are used as a sample for determining the corresponding correlations [22]. A linear relationship with a high correlation coefficient (greater than 0.999) is shown between the retention indices of the homologous series of alkylmethyl- and alkylethylfluorophosphonates, dimethylaminoalkyl-cyanophosphonate, and alkylmethylmethylphosphonate with the same structure of the alkyl radical [30].

It should be noted that the same class of substances could be compared with several rows of structural analogs, which provides independent control over the correctness of the obtained retention indices. Saturated alcohols can be considered as structural analogs of compounds containing an O-alkyl radical, because of their greatest availability.

Obviously, the method of structural analogy is based on the dependence of the properties of organic compounds on the number of identical structural units; therefore, many methods for evaluating the physicochemical properties of compounds are based on the use of additive models related to the conventional approaches in chemistry [31]. However, additive models become of little use in the complication of the structure of substances, because various kinds of intramolecular interactions and electronic effects appear. Apparently, the deviation of the experimental data from the calculated values in this case indicates the presence of these effects, and if they are correctly interpreted, the introduction of appropriate corrections minimizes discrepancies and makes possible the assessment of the correlations between the change in structure and property [32].

Since the regularities of the variation of most different physicochemical constants of organic com-

pounds (*A*) in any homologous series are identical, they can be described by the same elementary linear (first-order) recurrence relations [33−37], that is,

$$A(n + 1) = aA(n) + b, \qquad (6)$$

where *n* is the number of carbon atoms in the molecule, and *a* and *b* are coefficients (they are individual for one homologous series and are calculated from data for three or more simple homologues of the same series). Several computer programs enable calculating the boiling points, for example, according to the Clausius−Clapeyron equations and the Truton rule [38, 39]. To calculate the boiling point at atmospheric pressure by additive models, the ACD/Labs program [40] and computational bases on its basis [41] can be used. According to the data [42], the "structure−condition−property" model constructed on the basis of neural network simulation ensures a high accuracy of predicting the boiling points of hydrocarbons at different pressures in a wide range of experimental measurement conditions.

The main drawback of the above methods of calculation is their very approximate nature.

Application of fragment descriptors. Fragment descriptors (PDs) are a numerical characteristic of the chemical structure, indicating whether a specific structural fragment exists in it or specifying how many times it is contained in this structure [43]. Fragment descriptors are calculated using molecular fragments comprising certain atoms that play a specific role in the description of this characteristic [43−46]. The review [44] describes 11 main categories of fragment descriptors. The unique role of fragment descriptors is that they form a basis of the descriptor space; that is, any molecular descriptor (and any molecular property) that is an invariant of a molecular graph can be uniquely decomposed into this basis [47], which means that any structure−property relationship can be approximated using them.

Fragment descriptors can also be used as a basis for the methodology of additive models [48−50], based on the assumption of the total effect of the contributions of structural fragments (which are substituents attached to specific molecular compounds in a single series of compounds) or group contributions into the total value of the properties of the chemical compound.

Three main problems associated with the use of fragment descriptors are mentioned in publications [51]:

—a problem of "rare" or "missing" fragments;

—a problem of adequate representation of stereochemical information;

—absence of physical interpretation.

The problem of rare and missing fragments is, apparently, the most serious of the three mentioned. Indeed, the number of fragments and, consequently, the number of fragment descriptors are practically unlimited: they considerably exceed the number of possible chemical structures. As a result, any chemical structure contains fragments that are missing (or are present in too small amounts) in the training sample used to construct the QSAR/QSPR models needed to predict the desired property. Hence, a problem arises that is characteristic of the whole practice of application of fragment descriptors: the properties of organic compounds that are absent in the training sample cannot be reliably predicted [44, 52].

Application of quantum-chemical descriptors. Quantum-chemical descriptors (QCDs) [53] are numerical quantities obtained as a result of quantum chemical calculations. Such descriptors as the energy of the boundary molecular orbitals; partial charges at atoms and partial orders of bonds; the Fukui reactivity indices (free valence index, nucleophilic and electrophilic superdelocalizability); the energy of cationic, anionic, and radical localization; and dipole and higher multipole moments of the electrostatic potential are most often used as descriptors.

A variety of information on the structure of the test substance is usually obtained using computer programs, in particular, HYPERCHEM [54], GAUSSIAN [55], GAMESS [56], etc., which enable calculations by ab initio methods, semiempirical methods (PMX, CNDO, AM1, PM3), and other methods of molecular mechanics (MM+) [57]. Note that for all calculation methods of quantum chemistry, general limitations are typical caused by the impossibility of unambiguous selection of the most optimal ratio in parameters "accuracy" and "calculation rate" [58−60].

The success of the practical application of QCDs largely depends on the knowledge of the features of the approximations and the approach to parametrization of each method and the constraints imposed on the range of objects and properties that can be correctly calculated. For example, the standard deviation of the retention indices for 41 chlorinated dibenzo-*p*-dioxins, calculated by this method, was 25.16 index units, and the maximum deviation was 66.54 index units [61]. Such results are inferior to the requirements for the accuracy of analysis, which generally raises doubts about the advisability of using quantum-chemical descriptors for predicting retention indices.

Application of topological indices. Among the methods of mathematical simulation, the selection of which depends on the nature of the chemical compounds and the properties being analyzed, topological methods take a special place. Topological methods use only the information contained in the structural formula of the test compound [62] and suppose, first of all, the study of the interaction of atoms in a molecule [63, 64].

Topological indices (TIs) are some numerical parameters that are put in correspondence with the accepted image of molecules in the form of structural

Table 1. Topological indices [61, 70−76]

TI calculation equation	Drawbacks	Notes
Wiener index		
$W = \dfrac{1}{2}\sum\limits_{i=1}^{N}\sum\limits_{j=1}^{N} D_{ij},$ (7) where D_{ij} is the ith jth element of the distance matrix, which shows the shortest distance between the vertices i and j in graph G, that is, $D_{ij} = 1 = \dfrac{6}{z_i},$ (8) $D_{ij} = \sum \dfrac{1}{b}\dfrac{36}{z_i z_j},$ (9) where z_i and z_j are the charges of the nucleus (the number of all electrons) of atoms i and j connected by a given bond; b is a quantity characterizing the order (multiplicity) of the bond	Some of the information contained in the distance matrix is lost during the summation of its elements; therefore, the values of the Wiener index, referring to different isomers, are often overlapped	Proposed for describing the dependences of physicochemical properties of alkanes on their structure. Assumes the maximum value for unbranched isomers and decreases as their branching increases
Randic connectivity index		
$^{0}\chi = \sum (\delta_i)^{-\frac{1}{2}},$ (10) $^{1}\chi = \sum\limits_{g=1}^{k} (\delta_i \delta_j)^{-\frac{1}{2}},$ (11) where δ is the difference between the number of valence electrons and the number of hydrogen atoms associated with this atom; i and j are the numbers of atoms of the molecule formally connected by the bond under consideration; δ_i and δ_j are the δ values for two atoms of the bond; and k is the total number of bonds in the molecule	Does not distinguish between the m- and p-isomers of aromatic compounds	This index is a mathematically encoded information on the number of atoms in the molecule, their interconnections, and the degree of branching of the molecule; it can be calculated for different connectivity levels of atoms of the molecule with each other
Hosoya index		
$Z = \sum\limits_{k=0}^{\left[\frac{n}{2}\right]} p_k,$ (12) where p_k is the number of ways of selecting k edges of the graph so that no two of them have common vertices; n is the number of vertices of the graph	One of the hard-to-compute topological indices	It was used to assess the thermochemical properties of alkanes, bond orders, and π-electron energies in conjugated hydrocarbons

formulas [65, 66]. Topological indices are difficult to give a physicochemical interpretation, since their values are obtained through formal operations over graphs [67]. Currently, a number of varieties of topological indices are developed and applied: centric, composite, theoretical-information, topological-information, and topological-configuration [68]. Their values are determined not only by the structure, but also by the relative arrangement of the atoms of the compound under study on the plane [69]. The application of the so-called "three-dimensional indices" is described [61], in the calculation of which the real distances between atoms in the molecule are used. The most popular topological indices, widely used in theoretical chemistry, are given in Table 1 [61, 70−76].

A new semiempirical topological index (SETI) is presented in [77]. This index was originally developed to predict the chromatographic retention of linear and branched alkanes and alkenes [77, 78]. Later, it was applied to other classes of organic compounds: esters, ketones, aldehydes, and alcohols [79−85]. The SETI is calculated by equation

$$I_{ET} = \sum_i (C_i + \delta_i), \qquad (13)$$

where I_{ET} is a semiempirical topological index; C_i is the contribution of the value of each carbon atom or functional group in the molecule; δ_i is the sum of the logarithm of the value of each neighboring carbon atom and/or the logarithm of the value of the functional group.

Table 2. Results of the use of SETI for predicting the retention indices of classes of organic compounds on a standard non-polar stationary phase

Class of compounds	Number of isomers	Coefficients of equation $Ri = a + bI_{ET}$		Standard deviation	Reference
		a	b		
Alkanes	157	116.8000	−19.0500	26.2	[79]
Alkanes C_5–C_{14}	79	122.8446	−41.7054	2.3	[77]
Alkenes, ethers, alcohols	548	123.4758	−48.0866	7.0	[79]
Branched alkanes	59	120.4671	−29.0457	5.8	[78]
Methylbranched alkanes	178	123.1610	−39.5251	4.3	[78]
Esters	81	123.7900	−48.1400	5.8	[80]
Aldehydes and ketones	54	123.4951	−45.6553	5.0	[81]
Alcohols	44	124.1239	−51.3739	5.7	[82]
Alkylbenzenes	122	123.0632	−43.6579	8.3	[83]
Haloalkanes	141	124.7788	−56.8944	8.0	[84]
Aldehydes and ketones	31	123.8071	−47.5985	5.5	[85]
			Mean value	7.6	

SETI enables us to evaluate steric effects caused by the action of neighboring groups of substituents on the carbon atoms. Table 2 lists the results of the use of SETI for predicting the retention indices of organic compounds on a standard nonpolar stationary phase using a linear equation.

New topological indices are constantly developed. At present, more than 1000 topological indices are known, for which the dependence of physical and chemical properties on the structure of the compound has been searched. There are programs that allow you to select proper topological indices. For example, the DRAGON-7 program for describing the compound structure offers 5270 descriptors [86]. The selection of topological indices for solving the problem of finding the structure−property relation is, as a rule, quite random and largely depends on the intuition and ingenuity of the researcher. In this case, approximate formulas of different accuracy and reliability can be obtained for representing a property as a function of the same index. The most common criterion for assessing the suitability of a topological index is the square of the correlation coefficient, as well as the standard deviation, the Fisher test, etc. [87].

Some of the topological indices are related by simple relationships. In [88], the results of checking the approximate equations by equation

$$T_i = a + bT_j, \qquad (14)$$

are reported, where T_i and T_j are topological indices, and a and b are the coefficients of the linear equation. Coefficients a and b are selected by the least squares method. Such an analysis is performed for the topological indices of three series of hydrocarbons: alkanes, polyalkylbenzenes, and monocyclic structures. It is demonstrated [88] that between various topological indices, there are approximate linear relations with a sufficiently high correlation coefficient of $0.84 < r < 0.99$. At the same time, it should be noted that the indices of the same groups of compounds of different series contain approximately the same information about the structure, and, accordingly, they are characterized by common prediction errors.

Methods of mathematical statistics and machine learning. To solve the regression problem of structure−retention in the vector description of the structures of chemical compounds, the following methods of mathematical statistics and machine learning are most often used in chemoinformatics: multiple linear regression (MLR) [89]; partial least squares (PLS) [90]; artificial neural networks (ANN) [91], etc.

The histogram of the number of applications of mathematical statistics and machine learning for the simulation of chromatographic parameters is presented in Fig. 1. It is seen that the method of multiple linear regression is well studied and is most often used to predict the retention indices. Note that artificial neural networks became interesting to researchers only after a significant development of the method of backpropagation [92, 93]. Due to their ability to learn and generalize data, information retrieval systems built on the principle of an artificial neural network have been successfully applied in chemistry, especially in those cases where the analytical form of the relationship between the structure and properties of compounds is unknown [94, 95].

Table 3 shows the application of MLR for predicting the retention indices of organic compounds at standard weakly polar phases, published in the period from 2007 to 2013. Standard deviation is the main cri-
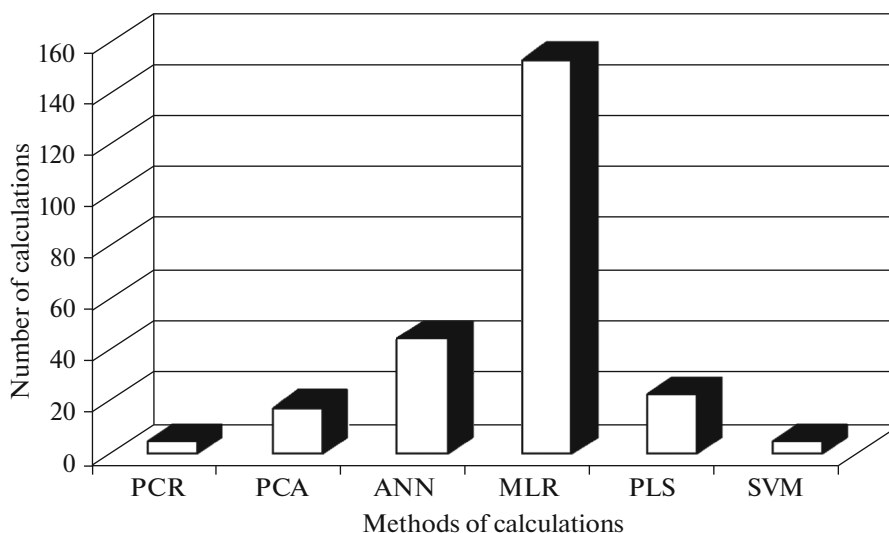
Fig. 1. Histogram of the number of applications of mathematical statistics and machine learning for the simulation of chromatographic parameters: MLR, multiple linear regression; ANN, artificial neural networks; PCR, principal component regression; PLS, partial least squares; PCA, principal component analysis; SVM, support vector machine.

terion for the accuracy of predicting the retention indices. The correlation coefficient of the methods considered below is from 0.9930 to 0.9999.

Table 4 compares the MLR, PLS, and ANN methods used to calculate the retention indices of organic compounds. The data presented in the table show that the accuracy of predicting the retention indices by these methods decreases in the series of ANN > PLS > MLR. In general, the total standard error for these methods is in the range from 7.6 to 151.0 index units. The average value of the standard deviation of all methods according to Tables 3 and 4 is 38.4 index units.

Thus, in the implementation of computational methods, data of a large sample of compounds with multiple branching of the carbon skeleton are used, and to improve the accuracy of the calculation of retention indices, descriptors are required that must be sensitive to structural isomerization, which is a significant limitation of this approach.

Use of additive models to predict the retention indices of organic compounds. In practice, because of the insufficient volume of experimental data on physicochemical characteristics that are not sensitive to the structural isomerization of topological and approximately calculated quantum-chemical descriptors, methods based on the construction of additive models according to the classical theory of the structure of molecules are used to calculate the retention indices of organic compounds.

A general approach to additive models for estimating the retention indices can be described by equation [3, 128]

$$Ri = Ri_0 + \sum_{i=1}^{N} \Delta Ri_i, \qquad (15)$$

where Ri_0 is the retention index of the "base" compound, which is taken as a basis for further transformations of the structure; ΔRi_i is the increment of the gas-chromatographic retention index corresponding to each of such structural transformations.

It is proposed to subdivide the additive models into three types, depending on the selection of Ri_0 and the method for calculating ΔRi_i [128]. The application of varieties of additive models is considered below. The additive models of the first type are based on the application of increments [129]. Increment ΔRi_i is numerically the difference between the retention indices of structures R–X and R–Y, that is, corresponds to a change in the retention index in the transformation of structure R–X into structure R–Y,

$$Ri = Ri_{R-X} + \Delta Ri_{(RX \to RY)}. \qquad (16)$$

The introduction of the same group of atoms into molecules of similarly constructed compounds changes the values of the retention indices by the same amount [129, 130].

There is a linear relationship between retention indices of homologues containing alkyl radicals with an unbranched carbon chain and the number of carbon atoms [21, 36]:

$$\log Ri = an + b, \qquad (17)$$

where a is a constant characterizing the contribution of the interaction energy of the methylene group with the stationary phase; b is a constant reflecting the contribution of energy of interaction of functional groups

Table 3. Results of using MLR for predicting the retention indices of organic compounds

Class of compounds (number of analytes)	Molecular descriptor	Standard deviation	Reference
Alkylpyridines (18)	Valence, molecular indices	27.7	[96]
Sulfides	New topological indices	12.6−27.3	[97]
Esters (90)	TD	8.1−18.2	[98]
Sulfur-containing heterocycles (114)	QCD	9.5	[99]
Solvents (49)	QCD	14.0−53.0	[100]
Various organic compounds (52)	Molecular weight, boiling point, McReynolds constant, branching index, etc.	7.0−26.7	[101]
Aldehydes (15), ketones (42)	Semiempirical electrotopological indices, SET	11.7	[102]
Volatile compounds (71)	Molecular weight, density, etc.	33.6	[103]
Esters (100)	Semiempirical electrotopological indices, SET, McReynolds constant	12.9	[104]
Polycyclic aromatic hydrocarbons (209)	Polarizability, Kier and Hali indices	12.8−13.4	[105]
Esters, ketones, aldehydes, alcohols (106)	Hydrogen-associations and electronegativity distance vectors	25.7	[106]
Alkylbenzenes (22)	Molecular weight, boiling point, the highest occupied molecular orbital occupied, the lowest unoccupied molecular orbital, dipole moment, etc.	17.5−26.8	[107]
Polycyclic aromatic sulfur-containing heterocycles (114)	Electronegativity distance vectors	9.5	[108]
Terpenes (47)	—	7.6	[109]
Polychlorinated dibenzofurans (115), polychlorinated dibenzodioxins (41), polychlorinated naphthalenes (62), polychlorinated biphenyls (210)	Generalized correlation index	17.9−40.6	[110]
Polychlorinated naphthalenes (62)	Electronegativity distance vectors	39.0	[111]
Substituted benzenes, benzaldehydes, and acetophenones (35)	Calculation of the polarity value for the stationary phase	20.4−22.8	[112]
Alcohols, ketones, ethers (106)	Modified molecular index of the polarizability effect, modifiable internal molecular polarizability index, TI	14.5−34.3	[113]
Alkyl-substituted cyclic hydrocarbons (174)	Charged local surface area of 56 descriptors	35.3	[114]
Hydrocarbons (134)	Descriptors calculated by the Dragon program (>400)	19.5−27.3	[115]
Various forensic structures (846)	Descriptors calculated by the Dragon program (526)	64.0−81.0	[116]
Terpenes	Topological indices	55.8	[117]
Various components of essential oils	—	74.9	[118]
Cyclic compounds of rosemary and essential oil of sage (40)	QCD, the lowest unoccupied molecular orbital, the highest occupied molecular orbital, polarizability, dipole moment, descriptors calculated by the Dragon program	51.9	[119]
Components of aromatic essential oils	One- and two-dimensional descriptors	58.1	[120]
Essential oils(116)	—	68.0	[121]
Polychlorinated biphenyls (209)	Descriptors calculated by the Dragon program	76.0	[122]
Methyl fatty esters (167)	Zero, one-dimensional, and two-dimensional descriptors	17.3−31.9	[123]
Pesticides (168)	—	151.0	[124]
Components of essential oils (100)	Descriptors calculated by the Dragon program (325)	50.3	[125]
Alkylphenols	The number of H atoms, the relative number of O atoms, the Balaban index	13.3	[126]
Aromatic components (656)	Topological and quantum chemical descriptors	59.6−61.0	[127]
	Mean value	34.7	

Table 4. Comparison of MLR, PLS, and ANN used for predicting the retention indices of organic compounds

Class of compounds (number of compounds)	Standard deviation			Reference
	MLR	PLS	ANN	
Various components of essential oils	74.9	88.8	48.9	[118]
Cyclic compounds of rosemary and essential oil of sage (40)	51.9	53.7	31.7	[119]
Components of aromatic essential oils	58.1	46.8–53.3	39.1	[120]
Compounds of essential oils (116)	68.0	57.0–61.0	49.0	[121]
Polychlorinated biphenyls (209)	76.0	35.0	63.0	[122]
Mean value	65.8	56.5	46.3	

with an adsorbent into the retention; and n is the number of carbon atoms.

Linear dependence (17) is more generally observed for nonpolar stationary phases, the retention at which is determined by the dispersion forces [26]; however, the first terms of the homological series do not obey the linear correlation. According to the results [26], this effect is because the properties of the methylene group located close to the functional group change under the influence of the latter. It is considered that the linear dependence (17) is observed, starting with the third or fourth member of the homologous series. In [26], examples of violations of the linear dependence for the fifth to eighth terms and above and various equations for taking into account the nonlinear change in retention indices as a function of the number of carbon atoms are given.

If the transformation leads to a more complex structure, then such a model of calculation is usually called a straight line. If the use of increments leads to a simplification of the structure, the model is called reversed [128]. The increments of similar transformations, for example, when a phenyl group is replaced by a methyl group, are averaged for different compounds and, as a result, reference values are obtained. The disadvantage of this approach is that when using increments, the structural features of specific compounds are not fully taken into account [128].

For example, the chromatographic retention indices of unknown compounds were calculated using incremental model in order to identify impurities contained in the reaction mixture formed during the chemical neutralization of O-isobutyl-S-(2-diethyl-aminoethyl)methylthiophosphonate [131, 132]. As structural analogies, the authors considered transitions from $-\text{N}(i\text{Pr})_2$ to $-\text{N}(\text{Et})_2$ and from O-ethyl to O-isobutyl substituents. The required increments of the retention indices of the corresponding structural transformations were estimated from the data obtained for simpler compounds. As a consequence, the average value of ΔRi in the case of recalculation of compounds with the $-\text{N}(i\text{Pr})_2$ moiety into compounds with the $-\text{N}(\text{Et})_2$ moiety was 146 ± 2 index units, and the mean value of ΔRi_i in the case of recal-

culation of compounds with the $(-\text{X}-\text{Et})$ moiety into compounds with the $(-\text{X}-i\text{Bu})$ moiety was 154 ± 7 index units. Using the obtained estimates of ΔRi, the authors successfully identified the decomposition products.

The increments of retention indices, obtained in the structural transformations of various functional groups of organic compounds were calculated [2, 133–137]. The variety of such increments is one of the main drawbacks hindering the wide application of additive models. Another problem of the effective use of ΔRi values is the inevitable loss of information about the structural features of molecules, which makes it difficult to control the correctness of their subsequent application [27].

The additive models of the second type differ from the previous ones by the absence of Ri_0; the desired index is calculated by adding atomic and group increments, and f is used as an additional correction:

$$Ri = \sum_{i=1}^{N} \Delta Ri_i + f, \qquad (18)$$

where f is the term the same as the average deviation of the calculated retention indices from the experimental values for compounds of the given series.

This model of calculation of the retention indices is used in the NIST program [138, 139]. The calculated retention index is given together with the confidence interval in index units; for example, for acetone, the confidence interval is 57 (50%) and 246 (95%) index units. This means that in the case of the 50% probability, the experimental retention index would be within 57 index units of the estimated retention index, and for the 95% probability, this limit would be 246 index units. Data analysis is based on erroneous measurements for the entire class of compounds (ketones) under study.

The additive model of the third type implements a modified or incrementless approach [128]. It is fundamentally different in that ΔRi_i, in fact, is calculated each time in the calculation of retention indices, which enables taking into account the structural features of compounds as precisely as possible. The mod-

Table 5. Number of possible isomers (*N*) of aliphatic alcohols and saturated hydrocarbons depending on the number of carbon atoms in the alkyl radical

Class of compounds	Number of carbon atoms										Total
	1	2	3	4	5	6	7	8	9	10	
Alcohols	1	1	2	4	8	17	39	89	211	507	879
Alkanes	1	1	1	2	3	5	9	18	35	75	150

ified additive model involves calculating the retention index based on data on the expected structures of the desired compounds by comparing the obtained values with the experimental ones. Such a version of the model can be called direct [128].

This version of additive models is based on the analogy of calculation methods with the approaches of hypothetical "assembly" of molecular structures of compounds to be characterized from simpler analogs [128, 140]. For example, if ABCD = ABC + BCD − BC, then *Ri*(ABCD) = *Ri*(ABC) + *Ri*(BCD) − *Ri*(BC). This method excludes the need for a preliminary calculation of increments and helps optimizing the selection of the initial structures by the criterion of their best fit for the compounds under study. It assumes the availability of reference values of the retention indices of selected precursors, and multistage assembly patterns should be avoided as this leads to a decrease in the reliability of the estimates obtained [128, 140].

The additive model of the third type was used to determine the structures of 839 congeners of polychlorinated hydroxybiphenyls [140], 211 structural isomers of 4-nonylphenols [141], cyclohexane chlori-

nation products [142, 143], condensation products of carbonyl compounds [144, 145], and chloro- and bromosubstituted anilines [146]. The regular increase in the number of isomers during the transition from the simplest to the higher homologues of any series with a significant decrease in the fraction of compounds characterized by different analytic parameters and physicochemical constants makes the use of a modified additive model particularly relevant, since it is optimal for calculating the retention indices of any series of compounds in the presence of data for the simplest homologues [130].

The maximum accuracy of the estimations of the retention indices can be achieved only when all the features of the structures of the target compounds that affect their gas-chromatographic retention parameters are reflected in the selected precursor structures [128].

The calculation of the retention indices by a modified additive model using saturated hydrocarbons shows that the same alkane structure can be used to predict two different types of branching compounds. Information on the number of structural isomers of aliphatic alcohols in comparison with saturated hydrocarbons (Table 5) can be used to calculate their retention indices. These data confirm the possibility of using one alkane to calculate the retention indices of several alcohols with different types of branching of the *O*-alkyl radical. Thus, the use of the retention indices of alkanes for predicting the retention indices of precursors can lead to serious errors.

In the analytical practice, there are modified additive models based on the indices of other compounds of the same series. In this case, the correlation dependences of the retention indices on the structure of the compounds are also used. The selection of equations

Table 6. Characteristics of prediction of the retention indices of organic compounds obtained using additive models

Class of compounds	Additive model	Prediction characteristics			Reference
		N	SD	Δ_{max}	
Polychlorinated dioxins	Modified by precursors	115	5.6−6.6	20−30	[150]
Chlorophenols	"	7	5.6	12	[140]
Chlorobromodioxins	"	63	3.3	7	[151]
Bromobenzofurans	"	115	6.3	19	
Chlorodioxins	"	42	5.6	12	[152]
4-Nonylphenols	Modified by alkanes	12	13.2	24	[141]
Alkylarenes	Incremental	21	6.3	24	[137]
Terpenes	"	49	20.2−31.7	53−73	[153]
Alkenes and cycloalkenes	"	26	6.7	20	[154]
Aromatic compounds	"	28	12.5	30	[154]
Alcohols	"	44	10.9	36	[154]
Aromatic compounds	"	92	25.7	153	[155]

N is the number of isomers; SD is the standard deviation; Δ_{max} is the maximum deviation from the reference value of the retention index.

describing such dependences is quite complex and unique for each series of compounds, which limits the universality of this method. Examples of this approach to the estimation of the retention indices of polychlorinated dibenzo-*p*-dioxins, dibenzofurans, and dibenzothiophenes are described [147−152].

Information on the application of additive models for different classes of organic compounds is summarized in Table 6. It is seen that the most accurate is the predicting of retention indices using a modified additive model, based on the application of data on the initial compounds of the class under consideration. In our opinion, this method takes into account all the structural features that affect the values of retention indices without taking into account additional criteria and corrections.

In general, many works are devoted to the development of methods for calculating and predicting the retention indices for capillary gas chromatography. Substantial progress is achieved in the framework of separate classes and series of organic compounds, which opens the possibility of creation in the near future of a universal software-computing device for obtaining the arrays of reference retention indices by calculation with accuracy acceptable for analytical practice.

REFERENCES

1. Trap, H.C. and Van Der Schans, M.J., *LC−GC Eur.*, 2007, vol. 20, no. 4, p. 202.

2. Zenkevich, I.G., *Mass-Spektrom.*, 2004, vol. 1, no. 1, p. 45.

3. Kovats, E., *Helv. Chim. Acta*, 1958, vol. 41, p. 1915.

4. Wehrli, A. and Kovats, E., *Helv. Chim. Acta*, 1959, vol. 42, p. 2709.

5. The official IUPAC website. http://www.iupac.org. Accessed January 2017.

6. Van den Dool, H. and Kratz, P.D., *J. Chromatogr. A*, 1963, vol. 11, p. 463.

7. *Identification of Potential Organophosphorus Warfare Agents: An Approach for the Standartization of Techniques and Reference Data*, Helsinki: The Ministry for Foreign Affairs of Finland, 1979, p. 202.

8. Manninen, A., Kuitunen, M.-L., and Julin, L., *J. Chromatogr. A*, 1987, vol. 394, p. 465.

9. Tsarev, N.I., Tsarev, V.I., and Kratkov, I.B., *Prakticheskaya gazovaya khromatografiya: Uchebno-metodicheskoe posobie dlya studentov khimicheskogo fakul'teta po spetskursu "Gazokhromatograficheskie metody analiza"* (Practical Gas Chromatography: A Manual for Students of the Chemistry Department on a Special Course "Gas Chromatography Methods"), Barnaul: Altaisk. Gos. Univ., 2000.

10. *High Resolution Gas Chromatography*, Hyver, K.J., Ed., Hewlett-Packard, 1989.

11. Zenkevich, I.G., *J. Anal. Chem.*, 2001, vol. 56, no. 9, p. 806.

12. Zenkevich, I.G., *J. Anal. Chem.*, 1998, vol. 53, no. 8, p. 725.

13. Automated Mass spectral Deconvolution and Identification System. http://chemdata.nist.gov/mass-spc/amdis/. Accessed January 2017.

14. Vershinin, V.I., Derendyaev, B.G., and Lebedev, K.S., *Komp'yuternaya identifikatsiya organicheskikh soedinenii* (Computer Identification of Organic Compounds), Moscow: Nauka, 2002.

15. The NIST 14 Mass Spectral Library with Search Program (NIST14/2014/EPA/NIH). Data Version: NIST 14. Software Version: 2.2g. www.nist.gov/srd/nist-standard-reference-database-1a-v14. Accessed January 2017.

16. Héberger, K., *J. Chromatogr. A*, 2007, vol. 1158, p. 273.

17. Héberger, K., in *Gas Chromatography*, Poole, C.F., Ed., Oxford: Elsevier, 2012, p. 451.

18. Raevsky, O.A., *Russ. Chem. Rev.*, 1999, vol. 68, no. 6, p. 505.

19. Todeschini, R. and Consonni, V., *Handbook of Molecular Descriptors*, New York: Wiley, 2008.

20. Zenkevich, I.G. and Kuznetsova, L.M., *Zh. Anal. Khim.*, 1992, vol. 47, no. 6, p. 982.

21. Zenkevich, I.G., *J. Struct. Chem.*, 1994, vol. 35, no. 6, p. 906.

22. Zenkevich, I.G., *J. Struct. Chem.*, 1999, vol. 40, no. 1, p. 101.

23. Panneerselvam, K., Antony, M.P., Srinivasa, T.G., and Vasudeva Rao, P.R., *Thermochim. Acta*, 2010, vol. 511, p. 107.

24. Prudkovskii, A.G. and Dolgonosov, A.M., *J. Anal. Chem.*, 2008, vol. 63, no. 9, p. 852.

25. Zenkevich, I.G., *Zh. Anal. Khim.*, 1995, vol. 50., no. 10, p. 1048.

26. Nabivach, V.M. and Dmitrikov, V.P., *Russ. Chem. Rev.*, 1993, vol. 62, no. 1, p. 23.

27. Zenkevich, I.G., *Russ. J. Gen. Chem.*, 2009, vol. 79, no. 10, p. 2164.

28. Grigor'ev, A.M., Bozhko, E.S., and Rudakova, L.V., *J. Anal. Chem.*, 2009, vol. 64, no. 2, p. 140.

29. Zenkevich, I.G., *J. Struct. Chem.*, 1996, vol. 37, no. 4, p. 674.

30. Soderstrom, M.T., Ketola, R.A., and Kostiainen, O., *Fresenius' J. Anal. Chem.*, 1995, vol. 352, p. 550.

31. Stein, S.E. and Brown, R.L., *J. Chem. Inf. Comput. Sci.*, 1994, vol. 34, p. 581.

32. Kurbatova, S.V., Finkelshtein, E.E., Kolosova, E.A., Kartashev, A.V., and Rashkin, S.V., *J. Struct. Chem.*, 2004, vol. 45, no. 1, p. 144.

33. Zenkevich, I.G., *Russ. J. Org. Chem.*, 2006, vol. 42, no. 1, p. 1.

34. Zenkevich, I.G., *Russ. J. Phys. Chem. A*, 2008, vol. 82, no. 6, p. 886.

35. Zenkevich, I.G., *Russ. J. Phys. Chem. A*, 2008, vol. 82, no. 5, p. 695.

36. Zenkevich, I.G., *Usp. Sovrem. Estestvozn.*, 2006, no. 7, p. 42.

37. Zenkevich, I.G., *Usp. Sovrem. Estestvozn.*, 2007, no. 4, p. 9.

38. Goodman, J.M., Kirby, P.D., and Haustedt, L.O., *Tetrahedron Lett.*, 2000, no. 41, p. 9879.

39. Estimation at required temperatures and pressures. http://calctool.org/CALS/chem/substance/boling. Accessed January 2017.

40. ACD/GC Simulator. ACD/Labs Version 6.00, build 6.06/19 Sep 2002.

41. The official Chemspider website. http://www.chemspider.com. Accessed January 2017.

42. Gal'bershtam, N.M., Baskin, I.I., Palyulin, V.A., and Zefirov, N.S., *Dokl. Chem.*, 2002, vol. 384, nos. 1−3, p. 140.

43. Zhokhova, N.I., Baskin, I.I., Palyulin, V.A., Zefirov, A.N., and Zefirov, N.S., *Dokl. Chem.*, 2007, vol. 417, no. 2, p. 282.

44. Zefirov, N.S. and Palyulin, V.A., *J. Chem. Inf. Comput. Sci.*, 2002, vol. 42, no. 5, p. 1112.

45. Baskin, I.I., Skvortsova, M.I., Stankevich, I.V., and Zefirov, N.S., Dokl. Akad. Nauk, 1994, vol. 339, no. 3, p. 346.

46. Baskin, I.I. and Varnek, A., *Comb. Chem. High Throughput Screening*, 2008, vol. 11, no. 8, p. 661.

47. Baskin, I.I., Skvortsova, M.I., Stankevich, I.V., and Zefirov, N.S., *J. Chem. Inf. Comput. Sci.*, 1995, vol. 35, no. 3, p. 527.

48. Bernstein, H.J., *J. Chem. Phys.*, 1952, vol. 20, no. 2, p. 263.

49. Laidler, K.J., *Can. J. Chem.*, 1956, vol. 34, p. 626.

50. Tetko, I.V., Bruneau, P., Mewes, H.-W., Rohrer, D.C., and Poda, G.I., *Drug Discovery Today*, 2006, vol. 11, nos. 15−16, p. 700.

51. Baskin, I.I., *Doctoral (Phys.−Math.) Dissertation*, Moscow: Moscow State Univ., 2010.

52. Baskin, I.I., Palyulin, V.A., and Zefirov, N.S., Abstracts of Papers, *Vsesoyuz. konf. po teoreticheskoi organicheskoi khimii* (All-Union Conf. on Theoretical Organic Chemistry), Volgograd, 1991, p. 557.

53. HyperChem 8.0.10. Trial version. http://www.hyper.com. Accessed January 2017.

54. The official Gaussian website. http://www.gaussian.com/index.htm. Accessed January 2017.

55. The official Gamess website. www.msg.chem.iastate.edu/gamess. Accessed January 2017.

56. Solov'ev, M.E. and Solov'ev, M.M., *Komp'yuternaya khimiya* (Computer Chemistry), Moscow: SOLON-Press, 2005.

57. Blatov, V.A. and Shevchenko, A.P., *Poluempiricheskie raschetnye metody kvantovoi khimii: uchebnoe posobie* (Semiempirical Computational Methods of Quantum Chemistry: A Manual), Samara: Univers-grupp, 2005, 2nd ed.

58. Barton, D. and Ollis, W.D., *Comprehensive Organic Chemistry. The Synthesis and Reactions of Organic Compounds*, Ed. Kochetkov, N.K., vol. 1: *Stereochemistry, Hydrocarbons, Halo Compounds, Oxygen Compounds*, Stoddart, J.F., Ed., Oxford: Pergamon, 1979. vol. 1

59. *Chemical Reactivity and Reaction Paths*, Klopman, G., Ed., New York: Wiley, 1974.

60. Kvant, S.D., Liu, H.X., and Wang, Z.Y., *QSAR Comb. Sci.*, 2007, vol. 26, p. 889.

61. *Chemical Applications of Topology and Graph Theory (Studies in Physical and Theoretical Chemistry)*, King, R.B., Ed., Amsterdam: Elsevier, 1983.

62. Seybold, P.G., May, M., and Bagal, U.A., *J. Chem. Educ.*, 1987, vol. 64, p. 575.

63. Makarov, L.I., *J. Struct. Chem.*, 1997, vol. 38, no. 4, p. 666.

64. Rouvray, D.H., Should we have designs on topological indices? in *Chemical Applications of Topology and Graph Theory (Studies in Physical and Theoretical Chemistry)*, King, R.B., Ed., Amsterdam: Elsevier, 1983, p. 159.

65. Balaban, A.T., *J. Chem. Inf. Comp. Sci.*, 1985, vol. 25, no. 3, p. 334.

66. Vinogradova, M.G., Papulov, Yu.G., Smolyakov, V.M., and Saltykova, M.N., *Russ. J. Phys. Chem. A*, 1996, vol. 70, no. 4, p. 627.

67. Randic, M., *J. Math. Chem.*, 1992, vol. 9, p. 97.

68. Nizhnii, S.V. and Epshtein, N.A., *Russ. Chem. Rev.*, 1978, vol. 47, no. 4, p. 383.

69. Malysheva, Yu.L., Papulov, Yu.G., Vinogradova, M.G., and Davydova, I.G., *J. Struct. Chem.*, 1998, vol. 39, no. 3, p 401.

70. Wiener, H., *J. Am. Chem. Soc.*, 1947, no. 69, p. 2636.

71. Wiener, H., *J. Am. Chem. Soc.*, 1947, no. 69, p. 17.

72. Dobrynin, A.A. and Mel'nikov, L.S., *Diskret. Anal. Issled. Operatsii, Ser. 2*, 2004, vol. 11, no. 2, p. 25.

73. Randic, M., *J. Am. Chem. Soc.*, 1975, no. 97, p. 6609.

74. Hosoya, H., *Bull. Chem. Soc. Jpn.*, 1971, no. 44, p. 2332.

75. Hosoya, H., *Internet Electron. J. Mol. Des.*, 2002, no. 1, p. 428.

76. Stankevich, M.I., Stankevich, I.V., and Zefirov, N.S., *Russ. Chem. Rev.*, 1988, vol. 57, no. 3, p. 191.

77. Heinzen, V.E., Soares, M.F., and Yunes, R.A., *J. Chromatogr. A*, 1999, vol. 849, p. 495.

78. Junkes, B.S., Amboni, R.D.M.C., Heinzen, V.E.F., and Yunes, R.A., *J. Chromatogr. A*, 2002, no. 55, p. 75.

79. Junkes, B.S., Amboni, R.D.M.C., Yunes, R.A., and Heinzen, V.E.F., *Internet Electron. J. Mol. Des.*, 2003, no. 2, p. 33.

80. Amboni, R.D.M.C., Junkes, B.S., Yunes, R.A., and Heinzen, V.E.F., *J. Mol. Struct.: THEOCHEM*, 2002, no. 579, p. 53.

81. Amboni, R.D.M.C., Junkes, B.S., Yunes, R.A., and Heinzen, V.E.F., *J. Mol. Struct.: THEOCHEM*, 2002, no. 586, p. 71.

82. Junkes, B.S., Amboni, R.D.M.C., Yunes, R.A., and Heinzen, V.E.F., *Anal. Chim. Acta*, 2003, no. 477, p. 29.

83. Porto, L.C., Souza, E.S., Junkes, B.S., Yunes, R.A., and Heinzen, V.E.F., *Talanta*, 2008, vol. 76, no. 2, p. 407.

84. Arruda, A.C., Junkes, B.S., Souza, E.S., Yunes, R.A., and Heinzen, V.E.F., *J. Chemom.*, 2008, vol. 22, nos. 3−4, p. 186.

85. Junkes, B.S., Amboni, R.M.C., Yunes, R.A., and Heinzen, V.E.F., *J. Braz. Chem. Soc.*, 2004, vol. 15, no. 2, p. 183.

86. The official website Dragon 7.0. https://chm.kodesolutions.net/products_dragon.php. Accessed January 2017.

87. Smolenskii, E.A., Slovokhotova, O.L., Chuvaeva, I.V., and Zefirov, N.S., *Dokl. Chem.*, 2004, vol. 397, nos. 4–6, p. 173.

88. Motoc, I., Balaban, A.T., Mekenyan, O., and Bonchev, D., *MATCH*, 1982, vol. 13, p. 369.

89. Draper, N.R. and Smith, H., *Applied Regression Analysis*, New York: Wiley, 1998, 3rd ed.

90. Linnik, Yu.V., *Metod naimen'shikh kvadratov i osnovy matematiko-statisticheskoi teorii obrabotki nablyudenii* (The Least Squares Method and the Foundations of the Mathematical-Statistical Theory of Processing Observations), Moscow: FIZMATLIT, 1958.

91. Haykin, S., *Neural Networks: A Comprehensive Foundation*, Upple Saddle River, NJ: Prentice Hall, 2001.

92. Rumelhart, D.E., Hinton, G.E., and Williams, R.J., Learning internal representations by error propagation, in *Parallel Distributed Processing*, Cambridge, MA: MIT Press, 1986, no. 1, p. 318.

93. Bartsev, S.I. and Okhonin, V.A., Adaptive information processing networks, *Preprint of Inst. Phys., Siberian Branch, USSR Acad. Sci.*, Krasnoyarsk, 1986, no. 59B, p. 20.

94. Gasteiger, J. and Zupan, J., *Angew. Chem., Int. Ed. Engl.*, 1993, vol. 105, no. 4, p. 503.

95. Devillers, J., *Neural Networks in QSAR and Drug Design*, London: Academic, 1996.

96. Tulasamma, P. and Reddy, K.S., *J. Mol. Graphics Modell.*, 2006, vol. 25, p. 507.

97. Zhang, T., Liang, Y.Z., Zhao, C.X., and Yuan, D.L., *Chin. J. Anal. Chem.*, 2006, vol. 34, p. 1607.

98. Liu, F., Liang, Y., Cao, C., and Zhou, N., *Talanta*, 2007, vol. 27, p. 1307.

99. Xu, H.Y., Zou, J.W., Jiang, Y.J., Hu, G.X., and Yu, Q.S., *J. Chromatogr. A*, 2008, nos. 1198–1199, p. 202.

100. Hoffmann, E.A. and Fekete, Z.A., Rajkó, R., Pálinkó, I., and Körtvélyesi, T., *J. Chromatogr. A*, 2009, vol. 1216, p. 2540.

101. Tello, A.M., Aguilar, R.L., Quintanilla-López, J.E., and Santiuste, J.M., *J. Chromatogr. A*, 2009, vol. 1216, p. 1630.

102. Souza, E.S., Kuhnen, C.A., Junkes, B.D.S., Yunes, R.A., and Heinzen, V.E.F., *J. Chemom.*, 2009, no. 23, p. 229.

103. Li, Z.G., Cao, H., Lee, M.-R., and Shen, D.-L., *J. Chromatogr. A*, 2009, no. 70, p. 1153.

104. Souza, E.S., Kuhnen, C.A., Junkes, B.D.S., Yunes, R.A., and Heinzen, V.E.F., *J. Mol. Graphics Modell.*, 2009, no. 28, p. 20.

105. Drosos, J.C., Viola-Rhenals, M., and Vivas-Reyes, R., *J. Chromatogr. A*, 2010, no. 1217, p. 4411.

106. Liao, L., Qing, D., Li, J., and Lei, G., *J. Mol. Struct.*, 2010, no. 975, p. 389.

107. Rykowska, I., Bielecki, P., and Wasiak, W., *J. Chromatogr. A*, 2010, no. 1217, p. 1971.

108. Li, Z.G., Cheng, F., Xia, Z., and Pu, S., *Acta. Chim. Sin.*, 2011, no. 29, p. 63.

109. Li, Z.G., Cao, H., Wang, L.L., Mo, W.M., and Shen, D.L., *Acta. Chim. Sin.*, 2009, no. 67, p. 289.

110. Peng, Z., Hu, M., Tian, F.F., and Li, Z.L., *Chin. J. Anal. Chem.*, 2006, no. 34, p. 1096.

111. Zhou, Y., Sun, L., Mei, H., and Li, S.Z., *J. Chromatogr. A*, 2006, no. 64, p. 565.

112. Ebrahimi, P. and Hadjmohammadi, M.R., *QSAR Comb. Sci.*, 2006, no. 25, p. 836.

113. Liu, F., Liang, Y., Cao, C., and Zhou, N., *Anal. Chim. Acta*, 2007, no. 54, p. 279.

114. Chen, H.F., *Anal. Chim. Acta*, 2008, no. 609, p. 24.

115. Zhang, X., Ding, L., Sun, Z., Song, L., and Sun, T., *J. Chromatogr. A*, 2009, no. 70, p. 511.

116. Garkani-Nejad, Z., *J. Chromatogr. A*, 2009, no. 70, p. 869.

117. Hemmateenejad, B., Javadnia, K., and Elyasi, M., *Anal. Chim. Acta*, 2007, no. 592, p. 72.

118. Noorizadeh, H., Farmany, A., and Khosravi, A., *J. Chin. Chem. Soc.*, 2010, no. 57, p. 982.

119. Noorizadeh, H. and Farmany, A., *J. Chromatogr. A*, 2010, no. 72, p. 563.

120. Noorizadeh, H. and Farmany, A., *J. Chin. Chem. Soc.*, 2010, no. 57, p. 1268.

121. Noorizadeh, H., Farmany, A., and Noorizadeh, M., *Quim. Nova*, 2011, no. 34, p. 242.

122. D'Archivio, A.A., Incani, A., and Ruggieri, F., *Anal. Bioanal. Chem.*, 2011, no. 309, p. 903.

123. Farkas, O., Zenkevich, I.G., Stout, F., Kalivas, J.H., and Héberger, K., *J. Chromatogr. A*, 2008, nos. 1198–1199, p. 188.

124. Jalali-Heravi, M., Ebrahimi-Najafabadi, H., and Khodabandehloo, A., *QSAR Comb. Sci.*, 2009, no. 28, p. 1432.

125. Riahia, S., Pourbasheer, E., Ganjali, M.R., and Norouzi, P., *J. Hazard. Mater.*, 2009, no. 166, p. 853.

126. Fatemi, M.H., Baher, E., and Ghorbanzade'h, M., *J. Sep. Sci.*, 2009, no. 32, p. 4133.

127. Yan, J., Cao, D.-S., Guo, F.-Q., Zhang, L.-X., He, M., Huang J.-H., Xu, Q.-S., and Liang, Y.-Z., *J. Chromatogr. A*, 2012, no. 1223, p. 118.

128. Ukolov, A.I. and Zenkevich, I.G., *Vestn. S.-Peterb. Univ., Ser. 4: Fiz., Khim.*, 2011, no. 1, p. 83.

129. Zenkevich, I.G. and Ukolov, A.I., *J. Struct. Chem.*, 2010, vol. 51, no. 4, p. 642.

130. Vigdergauz, M.S. and Semkin, V.I., *Zh. Fiz. Khim.*, 1972, vol. 45, no. 4, p. 948.

131. Savel'eva, E.I., Zenkevich, I.G., and Radilov, A.S., *J. Anal. Chem.*, 2003, vol. 58, no. 2, p. 114.

132. Savel'eva, E.I., Zenkevich, I.G., Kuznetsova, T.A., Radilov, A.S., and Pshenichnaya, G.V., *Ross. Khim. Zh.*, 2002, vol. 46, no. 6, p. 82.

133. Zenkevich, I.G., Eliseenkov, E.V., and Kasatochkin, A.N., *Mass-Spektrom.*, 2006, vol. 3, no. 2, p. 131.

134. Razgonyaev, A.O., Ukolov, A.I., and Zenkevich, I.G., *Analitika Kontrol'*, 2012, vol. 16, no. 3, p. 248.

135. Zenkevich, I.G., Eliseenkov, E.V., Kasatochkin, A., Zhakovskaya, Z., and Khoroshko, L., *J. Chromatogr. A*, 2011, no. 1218, p. 3291.

136. Zenkevich, I.G., Ukolov, A.I., Kushakova, A.S., and Gustyleva, L.K., *J. Anal. Chem.*, 2011, vol. 66, no. 12, p. 1165.

137. Zenkevich, I.G. and Chupalov, A.A., *Russ. J. Org. Chem.*, 1996, vol. 32, no. 5, p. 626.

138. Stein, S.E., Babushok, V.I., Brown, R.L., and Linstrom, P.J., *J. Chem. Inf. Model.*, 2007, vol. 47, p. 975.

139. The official website NIST 11 GC Method. Retention Index Database. http://chemdata.nist.gov/mass-spc/ri/. Accessed January 2017.

140. Zenkevich, I.G., Moeder, M., Koeller, G., and Schrader, S., *J. Chromatogr. A*, 2004, no. 1025, p. 227.

141. Zenkevich, I.G., Makarov, A.A., Shrader, S., and Moeder, M.A., *J. Chromatogr. A*, 2009, no. 1216, p. 4097.

142. Zenkevich, I.G., Eliseenkov, E.V., and Kasatochkin, A.N., *Mass-Spektrom.*, 2009, vol. 6, no. 2, p. 137.

143. Zenkevich, I.G., Eliseenkov, E.V., and Kasatochkin, A.N., *J. Chromatogr. A*, 2009, vol. 70, nos. 5−6, p. 839.

144. Zenkevich, I.G. and Ukolov, A.I., *Russ. J. Gen. Chem.*, 2011, vol. 81, no. 9, p. 1818.

145. Ukolov, A.I. and Zenkevich, I.G., *Mass-Spektrom.*, 2011, vol. 8, no. 4, p. 264.

146. Gruzdev, I.V., Alferova, M.V., Kondratenok, B.M., and Zenkevich, I.G., *J. Anal. Chem.*, 2011, vol. 66, no. 5, p. 504.

147. Katrizky, A.R., Maran, U., Lobanov, V.S., and Karelson, M., *J. Chem. Inf. Comput. Sci.*, 2000, no. 40, p. 1.

148. Liang, X., Wang, W., and Wu, W., *Chemosphere*, 2000, vol. 41, no. 6, p. 917.

149. Liang, X., Wang, W., and Wu, W., *Chemosphere*, 2000, vol. 41, no. 12, p. 1889.

150. Liang, X., Wang, W., and Wu, W., *Chemosphere*, 2000, vol. 41, no. 6, p. 923.

151. Donnelly, J.R., Munslow, W.D., Grange, A.H., Pettit, T.L., Simmons, R.D., and Sovocool, G.W., *J. Chromatogr. A*, 1991, no. 540, p. 293.

152. Donnelly, J.R. and Munslow, W.D., *J. Chromatogr. A*, 1987, no. 392, p. 51.

153. Duvenbeck, C.L. and Zinn, P., *J. Chem. Inf. Comput. Sci.*, 1993, no. 33, p. 220.

154. Peng, C.T., Ding, S.F., Hua, R.L., and Yang, Z.C., *J. Chromatogr. A*, 1988, no. 436, p. 137.

155. Acevedo-Martínez, J., Zenkevich, I.G., and Carrasco-Velar, R., *J. Chromatogr. A*, 2010, no. 71, p. 881.

Translated by O. Zhukova