

Linear Correlations of the Gas Chromatographic Retention Indices of Compounds from Various Taxonomic Groups

I. G. Zenkevich^a, *

^a Institute of Chemistry, St. Petersburg State University, St. Petersburg, 198504 Russia

*e-mail: izenkevich@yandex.ru

Received September 2, 2022; revised September 10, 2022; accepted September 10, 2022

Abstract—Despite the constant improvement of complex computer algorithms for the calculation of gas chromatographic retention indices (RIs), the simplest methods for their evaluation based on linear correlations between the indices of structural analogs from different taxonomic groups, $RI^1 \approx aRI^2 + b$, remain important.

It is shown that symbatic variations of the first numerical differences of retention indices, $\Delta_{RI}^1 = RI_{n+1} - RI_n$ (equivalent to the first derivatives of the retention indices with respect to the structural parameter varied in the group), are the conditions of correctness for such correlations in the simplest groups (substituted methanes). A monotonic variation of Δ_{RI}^1 in one of the groups with the presence of extrema in the other group is an unequivocal sign of the absence of a linear correlation between retention indices. If the values of Δ_{RI}^1 in one of the groups increase and decrease in the other, the ranking order of compounds in any one of them should be reversed. It is shown that the simplest relationship $RI^1 \approx aRI^2 + b$ is also applicable to more complex taxonomic groups (substituted ethanes, benzenes, and naphthalenes), and it allows one not only to estimate the RIs of compounds not yet characterized but also to refine known reference data.

Keywords: gas chromatographic retention indices, taxonomic groups of structural analogs, linear correlation, index correctness control, estimation of unknown values

DOI: 10.1134/S1061934823040160

The capabilities of identifying trace organic compounds in complex mixtures by gas chromatography–mass spectrometry can be effectively implemented only by the joint use of electron ionization mass spectra and gas chromatographic retention indices (RIs). However, judging by the contents of modern databases (including the NIST Mass Spectral Library [1]), the number of compounds characterized by mass spectra is significantly lower than the number of compounds for which RIs on standard stationary phases are known. For example, the NIST database [1] (version 2020) contains 350643 electron ionization mass spectra of 306869 compounds, but RI values are available for only 139498 compounds. Among them, both mass spectra and retention indices are known simultaneously for only a portion (111768) of test substances.

The limiting stage in the formation of databases of reference RI values is the availability of the test substances. Their target synthesis, if necessary, is the most time consuming step. For this reason, a rapid increase in the volume of reference data arrays is unlikely, and the improvement of methods for calculating retention indices based on molecular structures is of considerable current importance. There are a huge number of methods used for calculating RIs, the

consideration of which is an independent task. Recently, algorithms that involve the use of various topological descriptors (QSRR) have become popular [2–4], but self-learning algorithms based on the principles of neural networks [5–8] and more complex ones (for example, gene expression programming [9]) are becoming more widespread. At the same time, it cannot be argued that the possibilities of well-known and widely used additive schemes have been exhausted [10, 11].

The advantages of modern computer methods for calculating RIs [5–9] manifested themselves in the processing of the largest possible data samples, which is not always reasonable. The point is that, as new data are accumulated, the known reference values of RIs can be refined or even corrected (this also applies to the NIST database [1]), and this leads to the need to process the array of information for all compounds again. In addition, the accuracy of the estimates of retention indices in such algorithms is actually averaged over all of the test substances. This makes it difficult to identify less reliable data, which often has to be done manually. Moreover, the need to solve more local problems of estimating the RIs of single compounds using simpler methods is of current interest in

the practice of chromatographic analysis. At the same time, the accuracy characteristics of such estimates should relate to particular results rather than represent average values for the calculation method as a whole. In such cases, algorithms applicable to small taxonomic groups are effective, including the following:

(1) based on a correlation of RI with the position of a characterized compound in the corresponding group of homologs, congeners, or structural analogs, $RI = f(n_x)$;

(2) as a result of a correlation of the RI values of compounds from a selected taxonomic group with their physicochemical characteristics, $RI = f(A_x)$;

(3) as a result of a correlation of the RIs of compounds from a taxonomic group (1) with known values of the RIs of compounds from another group (2) (structural analogs), $RI^1 = f(RI^2)$.

Algorithms of the first group are an extended version of additive schemes; thus, their capabilities are comparable to each other. Methods of the second group provide the highest accuracy in estimating gas chromatographic retention indices, especially those involving the use of normal boiling points of homologs or congeners [12, 13]. The reason for this is that the number of compounds for which the values of T_b were determined with high accuracy significantly exceeds the number of compounds characterized by RIs. However, some complexity of these correlations lies in the fact that the relations used for calculations are nonlinear (linear logarithmic) [12, 13], which somewhat complicates the calculations:

$$\log RI = a \log T_b + bA + c, \quad (1)$$

where A is a parameter that characterizes the position of a particular compound in the corresponding taxonomic group (for example, the number of carbon atoms in a molecule, the number of rings, formal unsaturation, etc.); a , b , and c are coefficients calculated by the least squares method. The calculation of retention indices based on a correlation of their values for structural analogs of different groups (the third group; in the general case, such dependences are nonlinear) involves the use of similar linear logarithmic relationships [14]:

$$\log RI^1 = a \log RI^2 + bA + c, \quad (2)$$

where the meaning of the parameters A , a , b , and c is the same as in formula (1).

Considering the possibilities of correlation of gas chromatographic retention indices of structural analogs from different taxonomic groups, note a practically important special case. For some combinations of structural analogs from different classes, the function $RI^1 = f(RI^2)$ can be approximated with acceptable accuracy by the simplest linear regression equation

$$RI^1 \approx aRI^2 + b. \quad (3)$$

In such cases, the estimation of RIs is greatly simplified because it does not require the use of nonstandard calculation methods. However, no a priori rules for identifying groups of congeners to provide an acceptable accuracy of the results obtained by this variant have been formulated so far.

As an example, Fig. 1a illustrates the dependence of the retention indices of bromine-substituted methanes $CH_{4-x}Br_x$ with the numbers of bromine atoms from one to four ($1 \leq x \leq 4$) on the values of RIs for structurally similar chlorine-substituted methanes $CH_{4-x}Cl_x$ ($1 \leq x \leq 4$). It would seem that one should not expect large differences in the properties of these halogen derivatives; they are characterized by similar permittivities and dipole moments, and the atomic radii of bromine and chlorine are close (1.00 and 1.15 angstrom units, respectively). However, the function $RI(CH_{4-x}Br_x) = f[RI(CH_{4-x}Cl_x)]$ is nonlinear, and calculations under these conditions can be performed only using Eq. (2). At the same time, a similar relationship between the retention indices of iodomethanes and bromomethanes (Fig. 1b) is nearly linear (the correlation coefficient $R = 0.9989$), although the difference in the atomic radii of iodine and bromine is greater (1.40 and 1.15 angstrom units, respectively). Thus, from this example, it follows that the linearity or nonlinearity of the correlations $RI^1 = f(RI^2)$ is determined by some reasons other than the physicochemical characteristics of compared structural analogs.

Commenting on Fig. 1b, it should be noted that the RI of tetraiodomethane (1862) is disproportionately overestimated compared to those of other iodine derivatives: 516 ± 7 (CH_3I), 899 ± 9 (CH_2I_2), and 1209 ± 24 (CHI_3). It was included in the NIST database [1] based on two publications [15, 16], but, most likely, it was erroneous. The fact is that CI_4 melts at $168-171^\circ C$ with decomposition; therefore, its gas chromatographic analysis is impossible due to degradation in the chromatographic injector. Therefore, the RI value belonging to another compound (a decomposition product of CI_4) was replaced by an extrapolated theoretically expected estimate of 1446 calculated from linear logarithmic relation (2) based on data for the other three iodomethanes. As noted previously [17], checking the reliability of initial data is a necessary step of property-property correlations.

The purpose of this work was to identify criteria for the applicability of the simplest linear relations (3) to the estimation of gas chromatographic retention indices of congeners from various taxonomic groups based on the index values for congeners from other groups. The problems to be solved included the evaluation of unknown RI values, the verification of their correctness, and the identification of analytes in some cases.

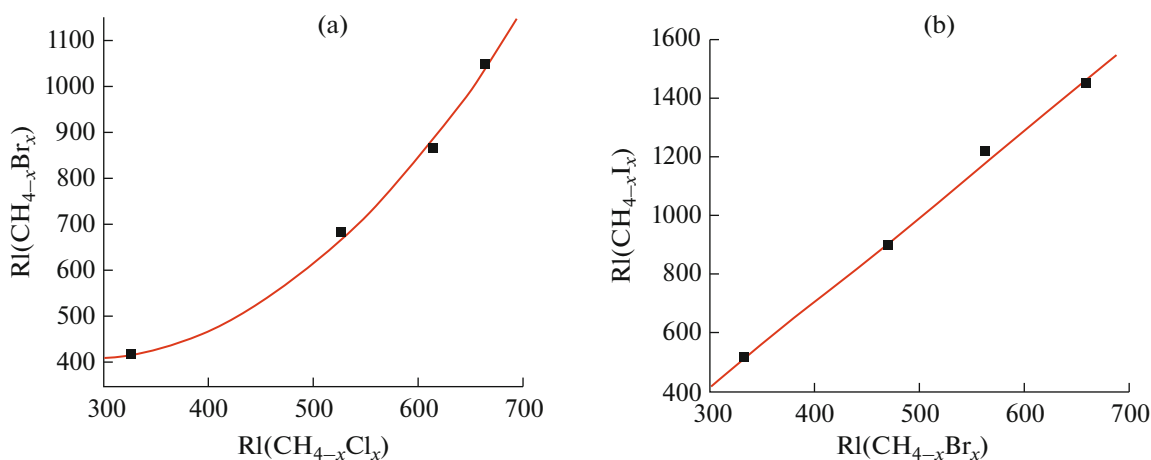


Fig. 1. Illustration of typical nonlinear and almost linear dependences of the retention indices of compounds from different taxonomic groups: (a) bromomethanes and chloromethanes, $1 \leq x \leq 4$; (b) iodomethanes and bromomethanes, $1 \leq x \leq 4$. (a) The nonlinear dependence can be linearized using Eq. (2). (b) Linear regression parameters: $a = 1.49 \pm 0.05$, $b = -101 \pm 39$, $R = 0.9989$, and $S_0 = 23$.

INITIAL DATA ACQUISITION AND PROCESSING (EXPERIMENTAL)

The NIST database [1] was a source of gas chromatographic retention indices of compounds under consideration on standard nonpolar polydimethylsiloxane stationary phases. The RI values of ethoxysilanol (C_2H_5O)_{4-x}Si(OH)_x ($1 \leq x \leq 3$) reported by Zenkevich and Baranov [18] were used (the conditions of gas chromatographic analysis were also specified).

Excel (Microsoft 2007) and Origin (versions 8.1 and 4.1) were used for least squares calculations. As criteria for the correctness of the linear relations $RI^1 \approx aRI^2 + b$, we selected (a) the correlation coefficients $R > 0.99$ and (b) the results of comparing the minimum RI differences of homologs from a characterized group with the generalized variances S_0 . The values of S_0 in the first approximation characterize the reliability of values calculated using a linear regression equation (for more accurate estimates, so-called error corridors are used [19]). The results of the approximation of RIs by relations (3) are significant if the minimum differences between the retention indices of neighboring structural analogs exceed twice the values of S_0 :

$$\min |RI_i - RI_{i\pm 1}| > 2S_0. \quad (4)$$

The failure of inequality (4) corresponds to unreliable estimates of RIs, which are inappropriate to use for solving analytical problems.

RESULTS AND DISCUSSION

Taxonomic groups of structural analogs. Features of linear correlations between the gas chromatographic retention indices of structural analogs were revealed using compounds from the following taxonomic types as examples:

(1) Substituted methanes: $CH_4 \rightarrow CH_3X \rightarrow CH_2X_2 \rightarrow CHX_3 \rightarrow CX_4$. Because the first member (CH_4 , RI = 100) in these groups is the same, it was excluded from consideration. As a result, only four compounds remained in them; this fact decreased the practical value of such RI correlations but simplified the identification of their regularities and features.

(2) Substituted ethanes: $C_2H_6 \rightarrow C_2H_5X \rightarrow CH_3CHX_2 \rightarrow XCH_2CH_2X \rightarrow XCH_2CHX_2 \rightarrow XCH_2CX_3 \rightarrow X_2CHCHX_2 \rightarrow X_2CHCX_3 \rightarrow X_3C-CX_3$. After the exclusion of unsubstituted ethane (RI = 200), eight compounds remained in these groups, and some of them (depending on X) were not characterized by RI values; because of this, the solution of such problems is a problem of greater current interest than that with compounds of the first type.

(3) Substituted benzenes: $C_6H_6 \rightarrow C_6H_5X \rightarrow C_6H_4X_2$ (3 isomers) $\rightarrow C_6H_3X_3$ (3 isomers) $\rightarrow C_6H_2X_4$ (3 isomers) $\rightarrow C_6HX_5 \rightarrow C_6X_6$. Without unsubstituted benzene (RI = 654 ± 7 [1]), the number of compounds in these groups was 13. For many of them, the estimation of RIs is of considerable current interest due to the lack of experimental data.

(4) Substituted naphthalenes: $C_{10}H_8 \rightarrow C_{10}H_7X$ (2 isomers) $\rightarrow C_{10}H_6X_2$ (10 isomers). These groups most often were limited to disubstituted compounds because reliable values of RIs are unknown for many isomers with even three the simplest substituents (methyl groups) [1].

Correlations of gas chromatographic retention indices in the group of substituted methanes and related compounds¹. Checking the mutual correlation of the

¹ Topological analogs of substituted methanes are, for example, substituted silanes.

Table 1. Retention indices of bromo- and chloromethanes [1] and their differences

Bromomethane	RI	Δ_{RI}^1	Chloromethane	RI	Δ_{RI}^1
CH ₃ Br	415 ± 5	268 181* 186	CH ₃ Cl	327 ± 3	201 87 48
CH ₂ Br ₂	683 ± 6		CH ₂ Cl ₂	528 ± 10	
CHBr ₃	864 ± 11		CHCl ₃	615 ± 6	
CBr ₄	1050 ± 5		CCl ₄	663 ± 16	

* Minimum value of Δ_{RI}^1 .

RI values of substituted methanes at different X (X = Cl, Br, I, OCH₃, OC₂H₅, CH₃, C₆H₅, etc.) makes it possible to identify examples of both nonlinear relations $RI^1 = f(RI^2)$ (Fig. 1a) and (more rarely) linear ones (as in Fig. 1b). To characterize the approach under consideration, it is necessary to elucidate the reasons for such differences, which, as noted above, are not related to the chemical nature of fragments X.

A necessary but not sufficient condition for the mutual linear correlation of the retention indices of structural analogs is their symbatic changes in both compared groups. If the values of RI monotonically increase or decrease in one of the groups and have an extremum in the second group, the application of relations (3) cannot provide the necessary accuracy of estimates. For example, the extremums (maximums) of retention indices were observed in the groups of methoxymethanes CH₃OCH₃ → ... → C(OCH₃)₄, fluoromethanes CH₃F → ... → CF₄, and methylsilanes CH₃SiH₃ → ... → Si(CH₃)₄.

To obtain more definite conclusions on variations in the values of RIs in different groups, it is necessary to compare not just the retention indices but their first numerical differences $\Delta_{RI}^1 = RI_{i+1} - RI_i$ (Table 1). Thus, for the above example of bromo- and chloromethanes (Fig. 1), the retention indices monotonically increased in either of the groups. However, the differences Δ_{RI}^1 in the series CH₃Cl → CCl₄ monotonically decreased from 201 to 48, while the last value of Δ_{RI}^1 (186) in the series CH₃Br → CBr₄ was somewhat larger than the previous one (181); that is, the set of the values of Δ_{RI}^1 was characterized by an insignificant extremum (minimum). In the mathematical sense, the first numerical differences Δ_{RI}^1 are equivalent to the first derivative of the function $RI = f(n_X)$ with respect to the number of variable fragments in the structure; however, it is a difficult task to identify the causes of variations in its sign (the appearance of extremums) even at a phenomenological level. Steric

interactions of fragments X in molecules, similar to the effects found previously [20] using methylbenzenes as an example, seem most likely.

Thus, a correlation of RI values cannot provide an acceptable accuracy of their estimates if the first differences of retention indices Δ_{RI}^1 change monotonically in one the compared taxonomic groups and are characterized by extremums in the second one. These examples include combinations of the data set of CH₃Cl → CCl₄ (the values of Δ_{RI}^1 decreased monotonically) with the data sets of CH₃OCH₃ → C(OCH₃)₄, CCl₃Br → CBr₄, and (CH₃)₃SiCl → SiCl₄ (in all cases, the values of Δ_{RI}^1 exhibited maximums).

On the other hand, if both of the compared sets of RI values are characterized by extremums of Δ_{RI}^1 , it can be assumed that their correlation is acceptable in terms of the accuracy of retention index estimates. Table 2 compares the retention indices of the series of methoxymethanes CH₃OCH₃ → C(OCH₃)₄ (data for which are not sufficiently reliable [1] and deserve clarification), considered as a function of the retention indices of chlorobromomethanes CCl₃Br → CBr₄. In both of the groups, the maximums of Δ_{RI}^1 were detected; however, the correlation of the RI values of these sets of structural analogs illustrated in Fig. 2 is suitable for obtaining practically useful index estimates. All values of Δ_{RI}^1 exceeded twice the generalized variance $S_0 = 23$. Table 2 summarizes the parameters of the linear regression equation and the calculated values of RIs; the average absolute deviation of calculated and reference values was only 15 ± 5 index units.

The values of Δ_{RI}^1 in two taxonomic groups to be compared should either increase or decrease simultaneously. Therefore, antibatic changes in Δ_{RI}^1 , which are equivalent to different signs of the first derivative dRI/dn_X , is another reason for the absence of the linear correlation $RI^1 = f(RI^2)$. A correlation of the retention indices of fluorochloromethanes CF_{4-x}Cl_x and

Table 2. Retention indices of chlorobromo- and methoxymethanes [1] and their differences

Methoxymethane	RI	Δ_{RI}^1	Approximated RI	Chlorobromomethane	RI	Δ_{RI}^1
CH ₃ OCH ₃	427 ± 19		408	CCl ₃ Br	757 ± 5	
		81				87
CH ₂ (OCH ₃) ₂	508 ± 2		531	CCl ₂ Br ₂	844 ± 1	
		194*				111*
CH(OCH ₃) ₃	702		689	CClBr ₃	955	
		128				95
C(OCH ₃) ₄	830 ± 4		824	CBr ₄	1050 ± 5	

* Maximum values of Δ_{RI}^1 .

chloromethanes CH_{4-x}Cl_x (1 ≤ x ≤ 4), the results of which are illustrated in Figs. 3a, 3b, can be given as an example. If both groups of structural analogs were formed by increasing the number of chlorine atoms in the molecules, that is, CF₃Cl → ... → CCl₄ and CH₃Cl → ... → CCl₄, the function RI¹ = f(RI²) was nonlinear (Fig. 3a). The reason is that the values of Δ_{RI}^1 increased in the former series and decreased in the latter series. To convert the data into a linear relationship, the sequence of structural analogs in the former (target) group should be reversed, that is, ranked by decreasing the values of Δ_{RI}^1 : CCl₄ → ... → CF₃Cl. However, as a result of this data transformation, a point (shown in Fig. 3b at the top left on an enlarged scale) corresponding to the pair of compounds CH₃Cl–CCl₄, which do not contain fluorine and, therefore, do not correspond to the other fluorine derivatives in chemical terms, appeared in the plot. Therefore, it is reasonable that it fell out of the general linear regression; the correlation coefficient for the three remaining pairs of compounds was R = 0.9995.

As an example of the application of the discussed correlation of RI values in the group of substituted methanes and their analogs (including silicon-containing ones), is interesting to consider the gas chromatographic identification of unusual products of the partial hydrolysis of tetraethoxysilane (C₂H₅O)₄Si (RI, 978 ± 10). Among them, triethoxysilanol (C₂H₅O)₃SiOH (RI, 727 ± 3), diethoxysilane diol (C₂H₅O)₂Si(OH)₂ (RI, 608 ± 1), and ethoxysilane triol (C₂H₅O)Si(OH)₃ (RI, 537 ± 2) were detected [18]. If we compared the RI values of these ethoxysilanols naturally ranked by increasing the number of ethoxy groups in the molecules (which corresponds to an increase in the values of Δ_{RI}^1) with the data for chloromethanes CH₃Cl → CCl₄ (which correspond to a decrease in Δ_{RI}^1), the function RI¹ = f(RI²) was found nonlinear (Fig. 4a). To linearize it, one of these sequences should be replaced by a reverse one, for

example, a set of ethoxysilanols should be considered based on an increase in the number of OH groups in the molecules: (C₂H₅O)₄Si → ... → (C₂H₅O)Si(OH)₃. As a result, we obtain an almost ideal linear regression (the correlation coefficient R is as high as 0.9995), the graph of which is shown in Fig. 4b. Note that, of course, the scales of axes in Figs. 4a and 4b are the same and special comments are required in order to distinguish between the two variants under consideration.

To additionally confirm the fact that the partial hydrolysis products of tetraethoxysilane formed a set of congeners, we can consider another correlation of their RI values with data for some other set of structural analogs, for example, the iodoalkanes Cl₄ (RI, 1446) → CHI₃ (1209 ± 24) → CH₂I₂ (899 ± 9) → CH₃I (516 ± 7). Bromoalkanes cannot be used for this purpose because of their inherent small maximum values of Δ_{RI}^1 (see Table 1). Figure 5 shows a plot of the

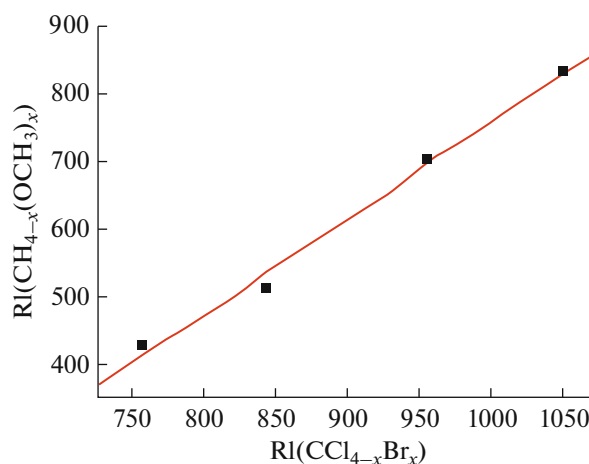


Fig. 2. Correlation of the retention indices of methoxymethanes and chlorobromomethanes (1 ≤ x ≤ 4). Linear regression parameters: a = 1.42 ± 0.10, b = -667 ± 95, R = 0.995, and S₀ = 23.

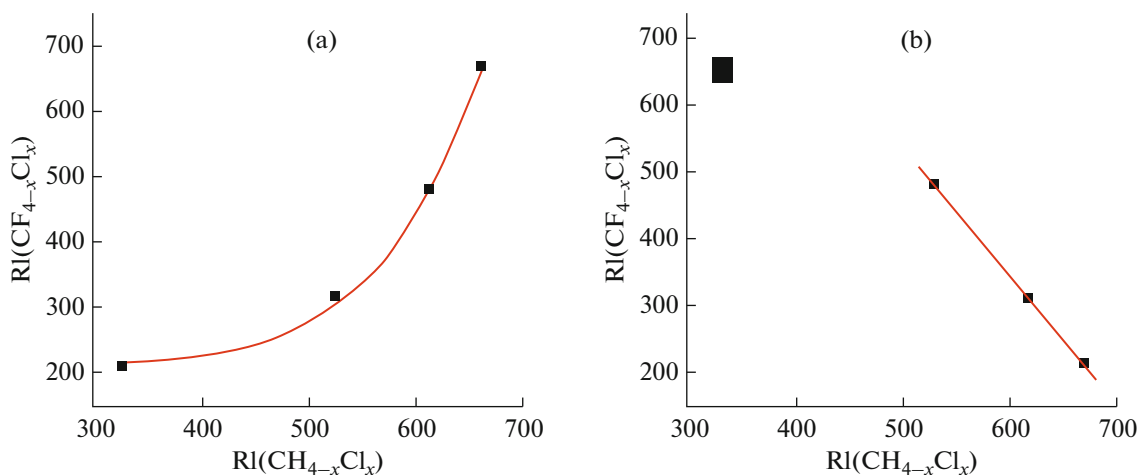


Fig. 3. Correlation of the retention indices of fluorochloromethanes and chloromethanes ($1 \leq x \leq 4$): (a) nonlinear dependence, both groups of structural analogs are ranked by increasing number of chlorine atoms in the molecules, $CF_3Cl \rightarrow CCl_4$ and $CH_3Cl \rightarrow CCl_4$; (b) groups of analogs are ranked by increasing values of Δ_{RI}^1 , $CF_3Cl \rightarrow CCl_4$, $CCl_4 \rightarrow CF_3Cl$, and $CH_3Cl \rightarrow CCl_4$. After excluding the $CH_3Cl-CCl_4$ point (upper left), the linear regression parameters for the three remaining pairs are $a = -2.00 \pm 0.06$, $b = 1536 \pm 38$, $R = 0.9995$, and $S_0 = 6$.

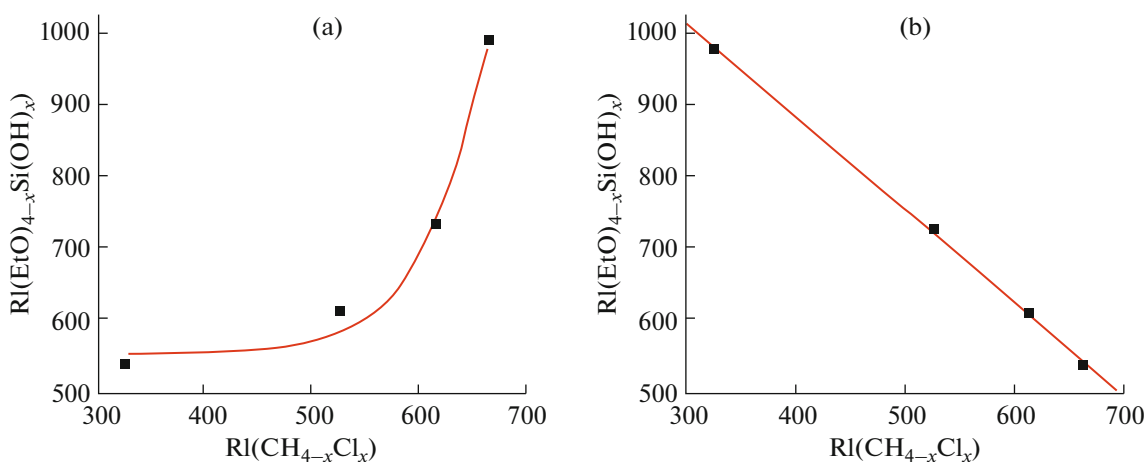


Fig. 4. Correlation of the retention indices of ethoxysilanols $(C_2H_5O)_{4-x}Si(OH)_x$ and chloromethanes ($1 \leq x \leq 4$): (a) nonlinear dependence, the RI values of ethoxysilanols are ranked by increasing molecular weights (or, otherwise, the number of ethoxy groups in the molecules), which corresponds to an increase in Δ_{RI}^1 ; (b) both groups are ranked by decreasing values of Δ_{RI}^1 , which leads to the reversal of the sequence of alkoxysilanols $(C_2H_5O)_4Si \rightarrow (C_2H_5O)Si(OH)_3$. (b) Regression parameters: $a = 1.30 \pm 0.03$, $b = 1408 \pm 16$, $R = -0.9995$, and $S_0 = 8$.

linear function $RI^1 \approx aRI^2 + b$. After eliminating a point corresponding to the pair of compounds $CH_3I-(C_2H_5O)_4Si$, the value of R for the remaining three pairs of structural analogs was -0.998 at $S_0 = 9$.

The considered correlations $RI^1 \approx aRI^2 + b$ are recommended to be used not only for evaluation but also for checking the reliability of reference retention indices of the least characterized “exotic” compounds. In this case, it is advisable to choose compounds with well-known and reproducible values of RIs as refer-

ence materials (used for comparison), for example, well-characterized hydrocarbons. The following example is of this type: it is necessary to check the known RI values of the methylsilanes $CH_3SiH_3 \rightarrow Si(CH_3)_4$. There are no data for the simplest monomethylsilane in the NIST database [1], and the next two homologs were characterized in a single work by Szekely et al. [21]; therefore, they have to be supplemented with calculated values based on the boiling points [12–14]. The RI values of methylsilanes can be

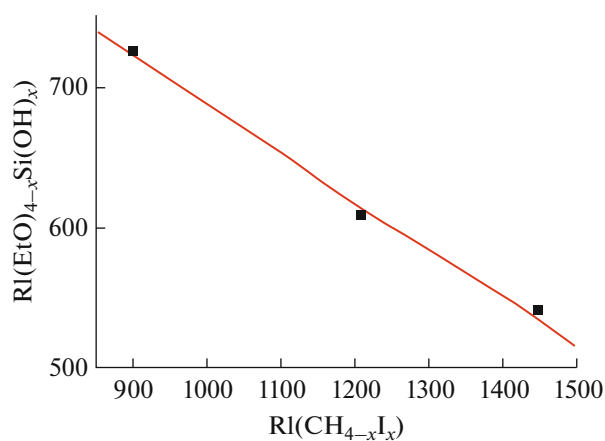


Fig. 5. Correlation between the retention indices of ethoxysilanols $(C_2H_5O)_{4-x}Si(OH)_x$ and iodomethanes ($2 \leq x \leq 4$). Both groups of structural analogs are ranked in descending order of Δ_{RI}^1 . Linear regression parameters: $a = -0.35 \pm 0.02$, $b = 1038 \pm 29$, $R = -0.998$, and $S_0 = 9$.

compared with the data for isostructural C_2 – C_5 alkanes, as illustrated in Table 3.

In this case, the graph of the function $RI^1 = f(RI^2)$ was very close to a linear one; thus, it is not shown here and only the following linear regression parameters are given: $a = 1.01 \pm 0.03$, $b = 39 \pm 9$, $R = 0.9992$, and $S_0 = 4.5$. Thus, it is confirmed that the known values of the RIs of methylsilanes are sufficiently reliable.

With an appropriate control of the variations Δ_{RI}^1 in the compared groups of structural analogs, the considered approach has no fundamental limitations, and it is characterized by exceptional simplicity.

Correlation of gas chromatographic retention indices in other taxonomic groups. This correlation (with a larger volume) significantly expands the capabilities of the approach under consideration. On the other hand,

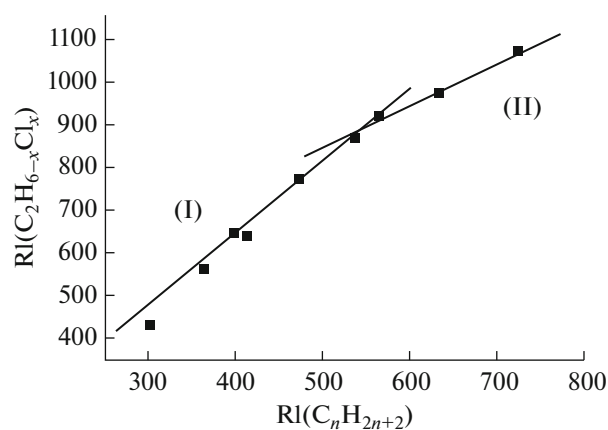


Fig. 6. Correlation between the retention indices of chloroethanes $C_2H_{6-x}Cl_x$ ($1 \leq x \leq 6$) and C_3 – C_8 alkanes. Linear regression parameters: $a = 1.88 \pm 0.08$, $b = 128 \pm 32$, $R = 0.997$, and $S_0 = 14$ for section (I); $a = 1.00 \pm 0.07$, $b = 339 \pm 43$, $R = 0.995$, and $S_0 = 10$ for section (II).

one can expect an unpredictable complication of the functions $RI^1 = f(RI^2)$, for example, due to the manifestation of difficult to formalize steric effects, which are noticeable only for individual compounds within these groups. As an example, it is expedient to consider a group of substituted ethanes (9 congeners), whose RI values are compared with the data for isostructural alkanes (Table 4). From their graphical representation (Fig. 6), it follows that the entire set of points is subdivided into two subgroups corresponding to two linear sections: (I) for compound nos. 1–6 (containing from one to four chlorine atoms) and (II) for compound nos. 6–9 (from four to six chlorine atoms). The caption to Fig. 6 specifies the parameters of two linear regression equations corresponding to each of these sections. At the moment, it seems difficult to explain the reasons for the appearance of two linear sections, and this fact should be treated as a phenomenologically revealed regularity. Similarly,

Table 3. Retention indices of methylsilanes and C_2 – C_5 alkanes [1]

Methylsilane	RI	Δ_{RI}^1	Approximated RI	Alkane	RI	Δ_{RI}^1
CH_3SiH_3	243 ± 8		240	C_2H_6	200	
		94				100
$(CH_3)_2SiH_2$	337 ± 8		342	C_3H_8	300	
		71				62
$(CH_3)_3SiH$	408		402	<i>iso</i> - C_4H_{10}	362 ± 7	
		48				51
$(CH_3)_4Si$	456 ± 16		413	<i>neo</i> - C_5H_{12}	413 ± 2	

Table 4. Retention indices of chloroethanes $C_2H_{6-x}Cl_x$ and C_3-C_8 alkanes [1]

Compound no.	Chloroethane	RI	Approximated RI (subgroup number)	Alkane (structural analog)	RI
1	C_2H_5Cl	422 ± 8	436 (I)	C_3H_8	300
2	$ClCH_2-CH_2Cl$	642 ± 7	624 (I)	C_4H_{10}	400
3	CH_3-CHCl_2	558 ± 19	553 (I)	Isobutene	362 ± 7
4	$ClCH_2-CHCl_2$	768 ± 4	761 (I)	Isopentane	473 ± 4
5	CH_3-CCl_3	641 ± 7	648 (I)	Neopentane	413 ± 2
6	$ClCH_2-CCl_3$	869	882 (I), 876 (II)	2,2-Dimethylbutane	537 ± 2
7	$Cl_2CH-CHCl_2$	916	905 (II)	2,3-Dimethylbutane	566 ± 3
8	C_2HCl_5	969	974 (II)	2,2,3-Trimethylbutane	635 ± 4
9	C_2Cl_6	1064	1063 (II)	2,2,3,3-Tetramethylbutane	724 ± 9

Table 5. Retention indices of bromobenzenes $C_6H_{6-x}Br_x$ and chlorobenzenes $C_6H_{6-x}Cl_x$ [1]

Chlorobenzene	RI	Bromobenzene	RI	Calculated RI value
Monochlorobenzene	839 ± 7	Monobromobenzene	839 ± 7	
1,2-Dichlorobenzene	1016 ± 11	1,2-Dibromobenzene	1216 ± 5	
1,3-Dichlorobenzene	989 ± 10	1,3-Dibromobenzene	1190 ± 7	
1,4-Dichlorobenzene	995 ± 10	1,4-Dibromobenzene	1193	
1,2,3-Trichlorobenzene	1199 ± 12	1,2,3-Tribromobenzene	—*	1530
1,2,4-Trichlorobenzene	1160 ± 13	1,2,4-Tribromobenzene	—	1464**
1,3,5-Trichlorobenzene	1117 ± 14	1,3,5-Tribromobenzene	1390	
1,2,3,4-Tetrachlorobenzene	1364 ± 11	1,2,3,4-Tetrabromobenzene	—	1808
1,2,3,5-Tetrachlorobenzene	1322 ± 11	1,2,3,5-Tetrabromobenzene	—	1737
1,2,4,5-Tetrachlorobenzene	1326 ± 18	1,2,4,5-Tetrabromobenzene	1697	
Pentachlorobenzene	1501 ± 5	Pentabromobenzene	—	2039
Hexachlorobenzene	1691 ± 5	Hexabromobenzene	2379	

* Dash means no reference data in the NIST database [1];

** RI is 1460, as estimated according to an additive scheme [10].

two linear portions of the function $RI^1 = f(RI^2)$ were observed when the RI values of chloro- and bromoethanes were correlated.

The average accuracy of the retention indices of chloroethanes estimated based on the data for isostructural hydrocarbons is 9 ± 5 index units, which is sufficient for practical purposes and not inferior to the capabilities of much more complicated methods. Because chlorine atoms can be located at two carbon

atoms, this fact complicates the use of the first numerical differences of the retention indices Δ_{RI}^1 (the array of Δ_{RI}^1 becomes two-dimensional).

It is advisable to start testing the capabilities of the considered approach in the groups of substituted benzenes by comparing the retention indices of bromo- and chlorobenzenes. Table 5 summarizes reference data; for some bromobenzenes, the RI values are

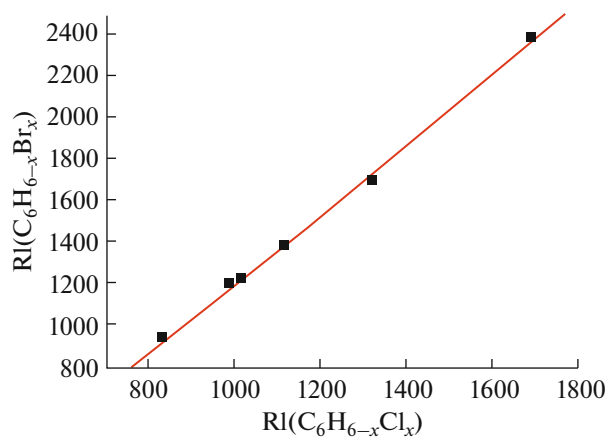


Fig. 7. Correlation between the retention indices of bromobenzenes $C_6H_{6-x}Br_x$ ($1 \leq x \leq 6$) and the corresponding chlorobenzenes. Linear regression parameters: $a = 1.68 \pm 0.03$, $b = -489 \pm 40$, $R = 0.9989$, and $S_0 = 24$.

unknown, and their assessment is of considerable current interest. Because the correlation coefficient R reached 0.9989, the plot of the function $RI^1 = f(RI^2)$ (Fig. 7) confirms its linearity.

An objective limitation of this method for assessing retention indices is the inability to distinguish between some position isomers characterized by relatively small differences in RIs, for example, *ortho*-, *meta*-, and *para*-dibromobenzenes with indices of 1216, 1190, and 1193, respectively. Their pairwise differences are lower than twice the value of $S_0 = 24$ (see Eq. (4)). It is noteworthy that there are no such coincidences for isomeric tribromobenzenes: 1530 (1,2,3-isomer), 1464 (1,2,4-isomer), and 1390 (1,3,5-isomer).

Similar estimates of RIs based on their mutual correlations for compounds from different taxonomic groups were obtained for the Br- and Cl-benzenes $C_6H_{6-x}Hal_x$ ($1 \leq x \leq 6$) as a result of their comparison with the data for methylbenzenes. It was effective to compare the retention indices of the iodobenzenes $C_6H_{6-x}I_x$ ($1 \leq x \leq 3$) and the methyl esters of benzenecarboxylic acids $C_6H_6-x(CO_2CH_3)_x$ ($1 \leq x \leq 4$) with the data for corresponding chlorobenzenes, the retention indices of the chloronaphthalenes $C_{10}H_{8-x}Cl_x$ ($1 \leq x \leq 4$) with the data for methylnaphthalenes, etc.

CONCLUSIONS

Thus, an analysis of the capabilities and features of the simplest linear correlation of gas chromatographic retention indices of compounds from different taxonomic groups of structural analogs, $RI^1 \approx aRI^2 + b$, made it possible to establish the following:

— If each compound in a particular group is uniquely defined by a molecular formula; that is, it has no isomers (this is possible for the simplest substituted

methanes, $CH_3X \rightarrow \dots \rightarrow CX_4$), systematic variations of the first numerical differences of the retention indices, $\Delta_{RI}^1 = RI_{n+1} - RI_n$, are responsible for a linear correlation of retention indices in these groups. The presence of extrema of Δ_{RI}^1 in one of the groups with a monotonic change in the other is an unambiguous sign of the absence of the linear correlation $RI^1 \approx aRI^2 + b$. If the values of Δ_{RI}^1 increase in one of the groups and decrease in the other, the ranking order of compounds should be reversed in any of the groups.

— It was shown that retention index correlations are possible in groups of structural analogs, such as substituted ethanes, benzenes, and naphthalenes. In all cases, the use of the simplest relation $RI^1 \approx aRI^2 + b$ makes it possible not only to estimate the values of RIs for compounds not yet characterized but also to refine unreliable reference data.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

REFERENCES

1. The NIST Mass Spectral Library (NIST/EPA/NIH EI MS Library, 2020 Release). Software/Data Version; NIST Standard Reference Database no. 69, Gaithersburg, MD: Natl. Inst. Standards Technol., 2020. <http://webbook.nist.gov>. Accessed September 2022.
2. Pompe, M. and Novic, M., *J. Chem. Inf. Comput. Sci.*, 1999, vol. 39, no. 1, p. 59. <https://doi.org/10.1021/ci980036z>
3. Farkas, O., Heberger, K., and Zenkevich, I.G., *Chemom. Intell. Lab. Syst.*, 2004, vol. 72, p. 173.
4. Heberger, K., *J. Chromatogr. A*, 2007, vol. 1158, p. 273. <https://doi.org/10.1016/j.chroma.2007.03.108>
5. Matyushin, D. and Buryak, A., *IEEE Access*, 2020, vol. 8, p. 223140. <https://doi.org/10.1109/ACCESS.2020.3045047>
6. Matyushin, D.D., Sholokhova, A.Yu., and Buryak, A.K., *Int. J. Mol. Sci.*, 2021, vol. 22, p. 13. <https://doi.org/10.3390/ijms22179194>
7. Idroes, R., Noviandy, T.R., Maulana, A., Syhendira, R., Sasmita, N.R., Muslem, M., Idroes, G.M., Kernala, P., and Irvanizam, I., *Int. Rev. Model. Simul.*, 2021, vol. 14, no. 2, p. 137. <https://doi.org/10.15866/iremos.v14i2.20460>
8. Qu, C., Schneider, B.I., Kearsley, A.J., Keyrouz, W., and Allison, T.C., *J. Chromatogr. A*, 2021, vol. 1646, p. 462100. <https://doi.org/10.1016/j.chroma.2021.462100>
9. Zhang, X., Shi, L., Ding, L., Sun, Z., Song, L., Qu, H., and Sun, T., *J. Chromatogr. Sep. Technol.*, 2015, vol. 6, no. 7, p. 7. <https://doi.org/10.4172/2157-7064.1000306>
10. Stein, S.E., Babushok, V.I., Brown, R.L., and Linstrom, P.J., *J. Chem. Inf. Model.*, 2007, vol. 47, p. 975. <https://doi.org/10.1021/ci600548y>

11. Zenkevich, I.G., Eliseenkov, E.V., and Kasatochkin, A.N., *Russ. J. Phys. Chem. A*, 2013, vol. 54, no. 3, p. 505.
<https://doi.org/10.1134/S0022476613030050>
12. Zenkevich, I.G., *J. Struct. Chem.*, 1994, vol. 35, no. 6, p. 906.
13. Zenkevich, I.G., *J. High Resolut. Chromatogr. Chromatogr. Commun.*, 1998, vol. 21, no. 10, p. 565.
14. Zenkevich, I.G., *J. Struct. Chem.*, 1996, vol. 37, no. 4, p. 674.
15. Waggott, A. and Davies, I.W., Identification of organic pollutants using linear temperature programmed retention indices (LTPRIs)—Part II, 1984. <http://dwi.defra.gov.uk/research/completed/research/reports/dwi0383.pdf>. Accessed August 2022.
16. Yasuhara, A., Morita, M., and Fuwa, K., *J. Chromatogr. A*, 1985, vol. 328, p. 35.
[https://doi.org/10.1016/S0021-9673\(01\)87374-X](https://doi.org/10.1016/S0021-9673(01)87374-X)
17. Zenkevich, I.G., *Russ. J. Phys. Chem. A*, 2021, vol. 95, no. 5, p. 894.
<https://doi.org/10.1134/S0036024421040294>
18. Zenkevich, I.G. and Baranov, D.A., *J. Anal. Chem.*, 2023, vol. 78, no. 1, p. 67.
19. Linnik, Yu.V., *Metod naimen'shikh kvadratov i osnovy teorii obrabotki nablyudenii* (The Least Squares Method and the Basics of the Theory of Processing Observations), Moscow: Fizmatgiz, 1958.
20. Zenkevich, I.G., *Russ. J. Phys. Chem. A*, 2021, vol. 95, p. 1720.
<https://doi.org/10.1134/S003602442107030X>
21. Szekeley, T., Nefedov, O.M., Gazzo, G., Shinyayav, V.I., and Fritz, D., *Acta Chem. Acad. Sci. Hung.*, 1967, vol. 54, nos. 3–4, p. 241.

Translated by V. Makhlyarchuk