*Journal of Structural Chemistry, Vol. 37, No. 4, 1996*

# STRUCTURAL ANALOGY PRINCIPLE FOR ESTIMATING GAS CHROMATOGRAPHIC RETENTION INDICES

**I. G. Zenkevich**

UDC 54.022:547.1:543.544

*It is shown that using additive schemes complying with the general principle of structural analogy for all molecular transformations is admissible for calculating gas chromatographic retention indices, which are very important analytical parameters for identification of organic compounds. From this viewpoint, we compare the possibilities of using the homomorphism factors and the fit increments of the retention indices of reagents and products in chemical reactions, including those of some topochemically equivalent hypothetical transformations of the structure and composition of organic molecules.*

Gas chromatographic retention indices (RI, $I_x$) are important one-dimensional analytical parameters for chromatographic and chromato-mass-spectrometric identification of organic compounds [1]. One of the reasons for their rare use is the absence of databases (of RI on standard phases) that would be comparable to modern mass-spectrometric databases in the number of defined compounds (up to 400,000 compounds in the mid-1990s). For this reason, the problem of calculating RI is not only of heuristic value; it is also important for expanding the available limited bases of reference data. Many methods were suggested for solving this problem [2]. However, their common limitations are unassessed accessibility and reliability of the initial data for calculating RI. As mentioned in [3], for this purpose it is not practical to use the parameters other than RI that are known for fewer substances; the same applies to quantities with low reproducibility or indirect relation to chromatographic retention parameters (for example, saturated vapor pressure, formation enthalpy and entropy, dipole moments, acidity and basicity functions, etc.). In this situation, most fruitful methods for calculating RI are those using the boiling points of organic compounds at atmospheric pressure, which are much more developed and systematized than those using indices themselves [4], or different additive schemes.

Calculation of RI from boiling points is based on very simple three-parameter correlation equations (1) and is possible not only for homologous series (molecular mass, molar refraction or the number of carbon atoms in the molecule serve as variable structural parameters $A$), but also for other taxonomic groups [5, 6]. The latter may be isologous ($A$ is formal unsaturation) or even genetic series ($A$ is the number of variable heteroatoms or functional groups in congeners):

$$\log I_x = a \log T_b + bA + c. \qquad (1)$$

Methods for calculating RI based on this equation provide the highest accuracy of this problem solving. Since the average errors $I_{exp} - I_{calc}$ do not exceed the average standard deviations of the randomized experimental values of RI (characteristic of the interlaboratory error of their determination), this actually means that it is not necessary to replace Eqs. (1) by more complex relations.

The wide use of additive schemes, which are less precise methods of evaluating RI, is explained by the absence of boiling points at atmospheric pressure for higher homologs of any series [a "natural" physicochemical limitation of the correlation dependences $I_x(T_b)$]. An optimal way to raise the reliability of these calculations seems to be combining several methods based on different principles. The standard errors for RI calculated by several methods are generally higher than those for randomized experimental data, but this does not hinder their use for identification of unknown compounds and addition to the corresponding databases.

---

Scientific-Research Institute of Chemistry, St. Petersburg University. Translated from *Zhurnal Strukturnoi Khimii*, Vol. 37, No. 4, pp. 784-795, July-August, 1996. Original article submitted April 24, 1995.

0022-4766/96/3704-0674$15.00

The efficiency of additive schemes for RI calculations is based on the additivity of sorbate–sorbent interaction energies, which is one of the main principles of chromatography [7]. However, construction of additive schemes obeys one more principle – structural analogy of substances within taxonomic groups. Homologous, isologous, and genetic series of compounds are formed in accordance with this principle. This paper compares different possibilities for calculating gas chromatographic RI on standard nonpolar phases using various composition and structure transformations of organic molecules that satisfy the above-mentioned principle.

## REQUIREMENTS TO INITIAL DATA FOR CALCULATIONS AND VERIFICATION OF RESULTS (EXPERIMENTAL)

Availability of databases for randomized indices on standard phases is an objective experimental background for verifying the applicability of the structural analogy principle for RI calculations. Currently, these databases include about 35,000 compounds for nonpolar polydimethylsiloxane elastomers. There are about 8000 organic compounds defined by RI on polar polyethyleneglycols, which is not enough to provide the necessary level of generalization.

Principles of RI randomization are stated in [3, 8]. The databases differ from the available databases (e.g., Sadtler Standard GC Retention Index Library) in standard deviations of indices ($S_I$), necessarily included in the former as parameters defining admissible discrepancies between experimental and reference data. The main reason for variations of $S_I$ is dissimilar temperature dependences of RI for compounds differing in the topochemical nature (primarily, with different numbers of rings). The known attempts to characterize the data by the total average values of $S_I$ (8-10 RI units for standard nonpolar phases) should be regarded as an initial approximation [9]. If the real values of $S_I$ are smaller than the average values, the use of the latter increases the ambiguity of identification; otherwise, the errors of identification of first kind increase in number.

Databases of randomized RI are designed for use in any laboratories and on any equipment provided that standard stationary phases are used. However, for additional control of correctness of results, addressing these databases should be preceded by testing the columns using specially designed test mixtures containing 10-15 components differing in chemical nature. Three "levels" of correction of results are possible depending on the average deviations of experimental values of $I_x$ from the randomized values of the database $\overline{\Delta I} \pm S_\Delta = \overline{(I_{\text{exp}} - I_{\text{rand}})}$:

| Relation between $S_\Delta$ and $\bar{S}_{\text{I, rand}}$ | Method of correction |
|---|---|
| $S_\Delta \le \bar{S}_{\text{I, rand}}$ | |
| 1. $\overline{\Delta I} \ll \bar{S}_{\text{I, rand}}$ | Correction is not required |
| 2. $\overline{\Delta I} \approx \bar{S}_{\text{I, rand}}$ | $I_x = I_{\text{exp}} - \overline{\Delta I}$; $S_{\text{I}} = S_{\text{I, rand}}$ |
| 3. $\overline{\Delta I} \gg \bar{S}_{\text{I, rand}}$ | $S_{\text{I}} = \sqrt{S^2_{\text{I, rand}} + \bar{S}^2_\Delta}$; $I_x = I_{\text{exp}} - \overline{\Delta I}$ |
| $S_\Delta > \bar{S}_{\text{I, rand}}$ | Correction is undesirable; databases for phases of different polarity must be used |

Cases 1 and 2 are typical for using packed and capillary columns with standard nonpolar polydimethylsiloxane phases SE-30, OV-1, OV-101, SP-2100, HP-1, DB-1, CP Sil 5 CB, etc. Correction of results (variant 3) permits the use of the suggested approach for interpreting the data obtained on more polar silicone phases SE-52, SE-54, OV-3, OV-7, HP-5, CP Sil 8 CB, etc. The same technique may be used for control of the calculated RI included in databases.

The increments of different transformations of organic structures discussed below were calculated from the corresponding values of RI for structural analogs with further statistical processing. The tendencies found may be extended, if necessary, to phases of different polarities.

## DISCUSSION

Simple calculations and analogy with the topochemical principles of organic chemistry are important advantages of using additive schemes for RI calculations. Among the numerous variants of such schemes, the best known variants are so-called $CH_2$ increments (approximately 100 RI units), substitution increments for benzene homologs ($I_y = I_{\text{RY}} -$

**TABLE 1.** Increments of Retention Indices of Some Structure Transformations of Alkyl Radicals A → B (homomorphism factors $H_x$)

| $\Delta M$ | A | B | $\overline{H} \pm S_{\mathrm{H}}(\mathrm{N})$ |
|---|---|---|---|
| 0 | $n\text{-}C_nH_{2n+1}$ | $-CH(CH_3)_2$ | -37±10(25) |
| | | $-CH(CH_3)CH_2CH_3$ | -29±12(9) |
| | | $-C(CH_3)_3$ | -77±14(14) |
| | | $-CH(CH_3)-CH(CH_3)_2$ | -41±7(6) |
| | | $-C(CH_3)_2CH_2CH_3$ | -62±6(7) |
| | | $-CH(C_2H_5)_2$ | -32±6(5) |
| | | $-CH(CH_3)-C(CH_3)_3$ | -75±12(6) |
| | | $-CH(CH_3)CH_2C(CH_3)_3$ | -114±9(5) |
| -2 | $n\text{-}C_nH_{2n+1}$ | $-CH=CH_2$ | -16±8(15) |
| | | $-CH=CH-CH_3$ *cis-* | 8±11(8) |
| | | *trans-* | 1±6(8) |
| | | $-CH=CH-C_2H_5$ | -3±5(6) |
| | $-CH(CH_3)_2$ | $-C(CH_3)=CH_2$ | -15±6(6) |
| | | $=C(CH_3)_2$ | 5±16(12) |
| | $-CH(CH_3)CH(CH_3)_2$ | $-C(CH_3)=C(CH_3)_2$ | 25±10(13) |
| -4 | $n\text{-}C_nH_{2n+1}$ | $-C\equiv CH$ | -2±5(10) |
| | | $-C\equiv C-CH_3$ | 53±7(8) |
| | | $-C\equiv C-C_2H_5$ | 27±3(6) |
| | | $-C\equiv C-C_mH_{2m+1}$ $(m\geq 3)$ | 19±3(6) |
| | | $-C=CH-CH=CH_2$ | 19±6(8) |
| | | $-CH=C=CH_2$ | 25±3(5) |
| +14 | $-CH_3$ | $-CHO$ | 182±8(22) |
| | $-C_2H_5$ | $-COCH_3$ | 168±10(20) |
| | $-C_3H_7$ | $-COC_2H_5$ | 171±8(8) |
| | $-C_nH_{2n+1}$ | $-COC_{n-1}H_{2n-1}$ | 150±12(4) |

$I_{\mathrm{RH}}$), and homomorphism factors (differences between the RI of the compounds under study and those of $n$-alkanes with the same number of carbon atoms in the molecule, $H_x = I_{\mathrm{RX}} - I_{\mathrm{RH}}$) [10]. However, the terms "homomorphism (isomorphism)" are polysemantic, and their usage is still vague; in many cases, these terms are equivalent to the notion of structural analogy, and even formal relations for $I_y$ and $H_x$ are identical. Although the above relations are well known, their practical use is limited because of the absence of statistically processed reference data.

Taking the above considerations into account, we calculated 25 RI increments of some very simple structural transformations of alkyl substituents, i.e., homomorphism factors $H_x$ as understood traditionally. These transformations are classified in Table 1 according to the parameter $\Delta M = M_{\mathrm{B}} - M_{\mathrm{A}}$ and include isomeric transformations of the hydrocarbon framework ($\Delta M = 0$), the introduction of multiple bonds ($\Delta M = -2$ and $-4$ for $C=C$ and $C\equiv C$, respectively), and the functionalization $CH_2 \rightarrow CO$ ($\Delta M = +14$). The list of these structural transformations may be considerably expanded if necessary.

Statistical reliability of the data of Table 1, neglected earlier, is of greater interest for us than applicability of this approach (which has long been known [10]). Standard deviations $S_H$ vary from 3 to 16 RI units, the average value being 8, which is comparable to recent estimates of the average interlaboratory error of RI determination on nonpolar phases [3].

Using RI increments corresponding to topochemically equivalent replacements of structural fragments or molecular compositions present another application of the structural analogy principle. This technique was developed in [11, 12] using the structurally equivalent classes of furans–thiophenes and 1,3-dioxolanes-1,3-dithiolanes as examples.

**TABLE 2.** Increments of Retention Indices of Some Topochemically Equivalent Transformations of the Structure and Composition of Organic Compounds A → B

| $\Delta M$ | A | B | $\overline{\Delta I} \pm S_{\bar{I}}(N)$ |
|---|---|---|---|
| I. Linear fragments | | | |
| +2 | $-CH_2$ | $-O-$ a* | –21±10(16) |
| | | b | 11±11(13) |
| +18 | $-CH_2-$ | $-S-$ | 195±13(25) |
| +1 | $-CH_3-$ | $-NH_2-$ | 123±22(14) |
| | $-CH_2-$ | $-NH-$ | 37±24(45) |
| | $>CH-$ | $>N-$ | 9±26(16) |
| +16 | $CR_4$ | $SiR_4$ | 14±13(11) |
| +16 | $-O-$ | $-S-$ | 191±11(8) |
| –1 | $-O-$ | $-NH-$ | 62±10(20) |
| | | c | 35±8(4) |
| –1 | $-CO-O-$ | $-CO-NH-$ | 270±17(14) |
| +16 | $-CO-O-$ | $-CO-S-$ | 154±10(13) |
| II. Alicyclic fragments | | | |
| +2 | $-CH_2-$ | $-O-$ (5, 6)** | 43±11(9) |
| | | (3, 4) | 74±21(13) |
| +18 | $-CH_2-$ | $-S-$ (5, 6) | 194±18(11) |
| | | (3, 4) | 231±19(8) |
| +1 | $-CH_2-$ | $-NH-$ (5, 6) | 87±18(10) |
| | | (3, 4) | 156±18(5) |
| +16 | $-O-$ | $-S-$ | 156±16(21) |
| –1 | $-O-$ | $-NH-$ | 57±18(12) |
| III. Aromatic fragments | | | |
| +16 | Furans (O) | Thiophenes (S) | 159±11(11) |
| | Benzofurans (O) | Benzothiophenes (S) | 213±15(3) |
| –1 | Furans (O) | Pyrroles (NH) | 217±11(8) |
| +1 | Arenes (CH) | Pyridines (N) | 99±15(12) |
| | | d | 45±26(16) |
| | Naphthalenes (CH) | Quinolines (N) | 56±15(20) |
| | | d | 0±21(19) |

---

*Additional structural constraints: a) substitution of a heteroatom for a carbon atom to obtain a $-C-X-C_mH_{2m+1}$ chain ($m \geq 2$); b) the same to obtain a $-C-X-CH_3$ chain; c) with two branchings of the carbon framework in the $\alpha$-position to the variable X fragment; d) with substituents in the 2 position of quinolines or pyridines.

**The value in parentheses is the size of the ring for which the given values of increments are valid.

Here again this approach, which is based on hypothetical (synthetically unfeasible) transformations of organic molecules, shows high generality and is not restricted to individual classes of compounds. Its realization also requires statistical treatment of both the initial data and the calculated values of $\Delta I$ for characterization by standard deviations.

Table 2 gives $\Delta I$ values for C → heteroatom and heteroatom → heteroatom structural transformations. Within each variety of these "substitutions," the values differ significantly between the linear and cyclic structural fragments (e.g., for O → S replacements, the values are 191±11 and 156±16, respectively). On the other hand, in alicyclic and aromatic systems the effects of O → S "substitutions" are almost equal (156±16 and 159±11, respectively). In some cases, additional structural restrictions need to be imposed, e.g., the size of the ring, the presence of substituents at the carbon atoms in the $\alpha$-positions to the variable structural fragments, and position of these fragments in the chain. The average standard deviations of RI increments for the transformations listed in Table 2 are twice as high as those in the previous case (16 RI units for variations from 8 to 26), showing that this group of calculation methods is slightly less reliable.

Another method of calculating RI of complex organic compounds, suggested recently, is based on the analogy with the schemes of synthesis of these compounds from simple reagents and uses increments characterizing organic reactions [13].

| Reaction scheme | Calculation of $\Delta I_R$ |
|---|---|
| I. A + ... → B + ... | $\Delta I_R = I_B - I_A$ |
| II. A + B → C + ... | $\Delta I_R = I_C - I_B - I_A$. |

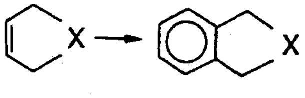
This approach is also a variety of additive schemes and reflects the general principle of structural analogy. Table 3 lists some values of $\Delta I_R$ for reaction schemes I and II. The potentialities of the method are discussed in detail in [13]; here for comparison with the above treatment it is worthwhile to pay attention primarily to the average values of standard deviations $S_R$, which are 18 [I] and 19 [II] RI units. According to this criterion, the RI increments of topochemically equivalent formal structural transformations (Table 2) and increments of real organic reactions (Table 3) are nearly equal in applicability.

Due to the wide range of structural transformations of molecules that can be of interest for RI calculations, one can suggest that further search for and expansion of this range has considerable promise. From this viewpoint, it is of interest to consider the previously unknown method of "functionalization" of hydrocarbons most completely defined by chromatographic parameters. A transition to less defined derivatives may be effected by "inserting" heteroatomic fragments into various positions of the carbon frameworks of molecules:

$$\text{C–C} \rightarrow \text{C–X–C}.$$

The major condition of the topochemical equivalence of these structural transformations is the necessity to take into account steric interactions between substituents at the neighboring carbon atoms, which can change when the substituents are separated by an X fragment. Table 4 characterizes only the cases $-CH_2-CH_2-$, $-CH_2-CHR-$, and (unless specified otherwise) $-CH_2-CR_2-$; therefore, the table is inapplicable to structures with two vicinal $-CHR-CHR-$ substituents, when the effects of two branchings of the carbon framework are nonadditive [14]. Also, other structural restrictions must be taken into consideration, for example, $p$–$\pi$ systems or $\pi$–$\pi$ conjugation of X fragments with CO and $C_6H_5$ groups or multiple bonds (usually characterized by different values of $\Delta I$). The $\Delta I$ parameters are nearly always dissimilar for the terminal C–C bonds when the formed structural unit contains the methyl group $-C-X-CH_3$ (the values are greater than those for $-C-X-C_nH_{2n+1}$ structures at $n \geq 2$). The introduction of a strongly polar carbamoyl group –CO–NH– into the carbon framework makes us provide for two more structural restrictions, associated with the presence of secondary and tertiary alkyl substituents at the nitrogen atom. However, given all additional conditions, in the latter case the standard deviations $S_{\Delta I}$ remain rather high (25-36 RI units). Due to this, it is preferable to calculate RI of amides using the O → NH substitution based on the data for esters (270±17, Table 2). If we disregard amides, the average standard deviation $\Delta I$ in Table 4 is 14 RI units, which is also comparable to the values of $S_{\Delta I}$ parameters given in Tables 2 and 3.

The various transformations of molecules that are consistent with the structural analogy principle may be illustrated by one more example (Table 4). The RI of polycyclic aromatic compounds may be estimated by comparisons with the corresponding alkenes, i.e.,



This method is convenient, e.g., for calculating RI of benzofurans, -thiophenes, or -pyrroles when experimental data are available only for the three simplest homologs of each series. The estimated standard deviations of this method are the same as those given above.

The methods of calculating RI discussed above and given in Tables 1-4 and other values of increments are

**TABLE 3.** Fit Increments of the Retention Indices of Reagents and Products in Some Organic Reactions (fragment of data array given in [13])

I. Transformation schemes A → C + ...

| $\Delta M = M_C - M_A$ | Reaction | $\overline{\Delta I}_R = \overline{(I_C - I_A)} \pm S_R(N)$ |
|---|---|---|
| –34 | $RBr \rightarrow RCH_2CH_2OH$ | 158±20(11) |
| –14 | $RR'CO \rightarrow RR'CH_2$<br>(Kishner–Wolff reaction) | –184±18(4) |
| –4 | $RCl \rightarrow ROCH_3$ | –23±7(5) |
| 0 | $RR'C{=}NOH \rightarrow RNHCOR'$<br>(Beckmann rearrangement) | 97±30(15) |
| 14 | $ArOH \rightarrow ArOCH_3$ | –62±16(9) |
| 34 | $ArH \rightarrow ArCl$ | 189±19(11) |
| 96 | $ROH \rightarrow ROCOCF_3$ | –85±24(7) |

II. Transformation schemes A + B → C + ...

| $\Delta M = M_C - M_A - M_B$ | Reaction | $\overline{\Delta I}_R = (I_C - I_A - I_B) \pm S_R(N)$ |
|---|---|---|
| –80 | $RC{\equiv}CH + RBr \rightarrow RC{\equiv}R'$ | –276±13(20) |
| –36 | $RCl + R'SH \rightarrow RSR'$ | –244±17(26) |
| –18 | $RCO_2H + R'OH \rightarrow RCO_2R'$ | –497±11(18) |
| –18 | $RCO_2H + R'R''NH \rightarrow RCONR'R''$ | –214±30(31) |
| –2 | $RSH + R'SH \rightarrow RSSR'$ | –94±15(13) |
| 2 | $ArNH_2 + ROH(+COCl_2) \rightarrow ArNHCO_2R$ | –14±8(12) |
|  | for *ortho*-substituted $ArNH_2$ | –57±7(4) |

needed not only for formation of RI databases. On the basis of RI additivity, new algorithms for interpreting chromato-mass-spectrometric data may be designed. For example, it is known that mass spectrometry is most efficient for group identification but does not give sufficient information on the structure of carbon frameworks for higher homologs [15]. Therefore, after any structural fragments of unknown compounds have been revealed by the characteristic signals of mass spectra, it is possible to subtract the RI increments of fragments from the experimental values of RI with subsequent chromatographic identification of simpler structures (in the limit, the problem may be reduced to identification of hydrocarbons) [16]. An example of using this approach is considered below (example 3).

The possibilities of different variants of RI calculation may be illustrated by the following examples.

1. Calculate RI of 2,2′-dichlorodiethyl sulfide (yperite) on standard nonpolar phases. The calculation is performed by one of several methods.

A. Method based on structural analogy with 2,2′-dichlorodiethyl ether (Chlorex) using the data of Table 2 for the O → S transformation:

$$(ClCH_2CH_2)_2O \rightarrow (ClCH_2CH_2)_2S,$$

$$954 \pm 7 + 191 \pm 18 = 1145 \pm 19.$$

B. Using the data for the possible reaction of synthesis of this compound, e.g., from 2-chloroethyl mercaptan and 1,2-dichloroethane [13]:

$$ClCH_2CH_2SH + ClCH_2CH_2Cl \rightarrow (ClCH_2CH_2)_2S$$

$$732 \pm 16 + 632 \pm 9 - 244 \pm 17 = 1120 \pm 25.$$

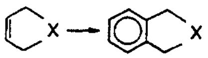
C. "Inserting" the sulfur atom into the carbon framework of 1,4-dichlorobutane:

$$Cl(CH_2)_4Cl \rightarrow (ClCH_2CH_2)_2S$$

$$874 \pm 5 + 272 \pm 13 = 1146 \pm 14.$$

The average of the three estimates is 1137±15; the randomized experimental value is 1131±10, i.e., the results coincide within the error. As mentioned above, the higher standard deviation of the calculated value compared to the interlaboratory scatter of RI is quite normal.

**TABLE 4.** Increments of Retention Indices of Some Structural Fragments Introduced into the Carbon Framework of Organic Molecules C–C → C–X–C

| $\Delta M$ | –X– | Additional structural constraints | $\overline{\Delta I} \pm S_I$ |
|---|---|---|---|
| 15 | NH | a | 132±14(5) |
| | | b | 172±9(10) |
| 16 | O | a | 80±8(10) |
| | | b | 115±10(10) |
| | | Conjugation with $C_6H_5$ | 123±12(11) |
| | | *ortho*-Substituted $C_6H_5$ | 75±75±12(13) |
| | | Conjugation with C=C a | 90±13(9) |
| | | b | 115±12(3) |
| 24 | C≡C | a | 225±5(7) |
| | | b | 252±7(7) |
| 26 | CH=CH (*trans*) | | 202±6(13) |
| | | Conjugation with $C_6H_5$ | 258±14(10) |
| | | Conjugation with C=O | 261±7(6) |
| | | Conjugation with C=C | 237±9(6) |
| 28 | CO | | 265±13(23) |
| | | Conjugation with $C_6H_5$ | 285±17(19) |
| | | *ortho*-Substituted $C_6H_5$ | 232±14(5) |
| | | Conjugation with C=C | 296±19(7) |
| 28 | N=N | | 158±15(13) |
| 32 | S | a | 272±13(17) |
| | | b | 300±9(12) |
| 43 | CO–NH | a | 532±25(21) |
| | | b | 589±33(7) |
| | | c | 499±33(6) |
| | | d | 447±36(3) |
| 44 | CO–O | a | 268±13(31) |
| | | b | 299±11(8) |
| | | e | 232±9(4) |
| 56 | CO–CO | | 368±10(7) |
| 60 | CO–S | | 446±22(12) |
| 64 | S–S | | 504±20(12) |
| 74 | $Si(CH_3)_2$–O | | 285±15(21) |
| 76 | *n*-Phenylene-($C_6H_4$) | a | 627±14(8) |
| | | b | 644±12(8) |
| | *o*-Phenylene | a | 621±29(7) |
| | | b | 663±21(8) |
| 50 | |  | |
| | | Arenes → naphthalenes | 527±18(20) |
| | | Furans → benzofurans | 470±13(3) |
| | | Thiophenes → benzothiophenes | 530±9(3) |
| | | Pyrroles → benzopyrroles | 570±3(3) |
| | | Cycloalkenes → benzocycloalkanes | 464±15(6) |

Note. Additional structure limitations: a) X fragment introduced into any positions of the carbon chain other than terminal; b) increments for the C–$CH_3$ terminal bond; c) secondary radical at nitrogen; d) tertiary radical at nitrogen; e) two or more branchings of the carbon framework in the $\alpha$-positions relative to the X fragment.

2. Estimate the currently unknown RI of dimethyl 1,4-phenylene diacetate $CH_3OCO{-}CH_2{-}C_6H_4{-}CH_2{-}CO_2CH_3$.

The calculation according to the scheme in Table 2 is impossible in this case, because data on structural analogs are lacking. This is just the situation that clearly shows advantages of using a set of calculation methods. Thus the scheme using the analogy with possible methods of synthesis of this ester is more universal and applicable for estimating RI even if it is not optimal by chemical criteria.

A. Synthesis of the compound is theoretically possible using the Friedel–Crafts reaction from methyl phenylacetate (RI 1150±10) and methyl bromoacetate (801±10); the reaction increment is −321±11 [13]:

$$1150\pm10 + 801\pm10 - 321\pm11 = 1630\pm18.$$

B. Another method uses a multistep reaction and is therefore less accurate. The initial data are the RI of 1,4-dipropylbenzene (1218±5), the doubled increment of the Kishner–Wolff reaction with an opposite sign (the scheme involves the reaction of transformation of 1,4-dipropylbenzene into 1,4-phenylene-bis(2-propanone), which is hypothetically reverse to the real reaction), and the doubled increment of the Baeyer–Villiger reaction $RCOCH_3 \rightarrow RCO_2CH_3$ (17±17) [13]:

$$1218\pm5 + 2\cdot(184\pm18) + 2\cdot(17\pm17) = 1620\pm35.$$

C. Insertion of two –CO–O– fragments into the carbon framework of 1,4-diethylbenzene (RI 1040±8); this is one of the simplest calculation methods:

$$1040\pm8 + 2\cdot(299\pm12) = 1638\pm18.$$

D. Introduction of the *para*-phenylene fragment into the dimethyl succinate molecule (RI 1006±15):

$$1006\pm15 + 644\pm12 = 1650\pm19.$$

Thus, as in example 1, all calculated values of RI may be considered equivalent, since they coincide within the errors. The final result is obtained by averaging these values: 1634±13.

3. The last example illustrates the "reverse" logic of using the increments of structural transformations of molecules for the purpose of refinement of unknown structures by experimental chromato-mass-spectrometric data in solving practical identification problems but not for RI calculations.

Derive the most detailed information about the structure of an unknown ester (homologous series – alkylalkanoates – established from the spectrum of the ionic series [15]) with the retention index 902±2, determined on an OV-101 capillary column, and with the following mass spectrum [8 fundamental signals, $m/z$ ($I_{\mathrm{rel}}$)]: 43(100), 71(87), 41(52), 56(42), 57(41), 27(40), 29(31), 89(22). No signals of molecular ions have been recorded.

Using the data of Table 4, we can estimate RI of the alkane structural analog of this ester after the –CO–O– fragment has been "eliminated" from the carbon framework of the molecule. However, as no *a priori* data on the nature of this ester are available, in the first step one needs to provide for three variants of data analysis:

$$I_x = 902\pm2$$

| a. No more than one branching of the carbon framework in the $\alpha$-position to the –CO–O– fragment | b. Methyl ester | c. Two or more branchings of the carbon framework in the $\alpha$-position to the –CO–O– fragment |
|---|---|---|
| $\Delta I = -268\pm13$ | $\Delta I = -299\pm11$ | $\Delta I = -232\pm9$ |

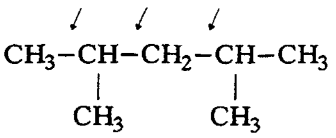
Consequently, possible ranges of RI of the hydrocarbon analogs are

| 634±13 | 603±11 | 607±9. |
|---|---|---|

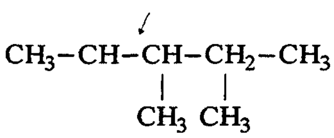
These ranges correspond to the following alkane structures established by searching in chromatographic databases:

| | | |
|---|---|---|
| 1. 2,2-Dimethylpentane (627±2) | 4. *n*-Hexane (600) | 5. 2-Methylhexane (667±1) |
| 2. 2,4-Dimethylpentane (630±2) | | 6. 3-Methylhexane (676±1) |
| 3. 2,3,3-Trimethylbutane (640±4) | | 7. 2,3-Dimethylpentane (672±2) |

However, after analyzing the initial and final information for compatibility (logical consistency), we can considerably reduce the list of seven variants of the structure of the carbon framework. For example, methyl ester [4] must be excluded from consideration, because there is no signal with $m/z = 74$ in the mass spectrum. Variants 5 and 6 are eliminated because of the absence of two or more branchings of the carbon framework. Structures 1 and 3 should also be discarded for branching reasons and the ban on methyl esters. As a result, we have only two variants (the arrows show the C–C bonds in which the –CO–O– fragment may be "inserted"):



$$\mathrm{CH_3\overset{\downarrow}{-}\underset{\displaystyle CH_3}{\underset{|}{CH}}\overset{\downarrow}{-}CH_2\overset{\downarrow}{-}\underset{\displaystyle CH_3}{\underset{|}{CH}}-CH_3}$$



$$\mathrm{CH_3-\underset{\displaystyle CH_3}{\underset{|}{CH}}\overset{\downarrow}{-}\underset{\displaystyle CH_3}{\underset{|}{CH}}-CH_2-CH_3}$$

Further reduction of the list again demands addressing the mass spectrum, in which the fundamental signals with $m/z = 71$ and 43 belong to the acyl fragment of the molecule:

$$[C_3H_7CO]^+ \ (m/z = 71) \rightarrow [C_3H_7]^+ \ (m/z = 43).$$

Hence the above set of analytical parameters is satisfied by isobutyl isobutyrate (experimental value of RI is 902±3) and *sec*-butyl isobutyrate (RI is 900). Any other variants are ruled out. We note that this result was obtained without information on the molecular mass of the unknown compound.

This work was supported by the International Science Foundation (grant No. NX 4000).

## REFERENCES

1. F. W. Karaseek and R. E. Klement, *Basic Gas Chromatography-Mass Spectrometry*, Elsevier, Amsterdam (1988).
2. R. Kaliszan, *Quantitative Structure-Chromatographic Retention Relationships*, Wiley, New York (1987).
3. I. G. Zenkevich and L. M. Kuznetsova, *Collect. Czech. Chem. Commun.*, **56**, No. 10, 2042-2056 (1991).
4. *Beilstein Handbuch der Organischen Chemie*, Vols. I-XXXI, Berlin (1918).
5. I. G. Zenkevich, *Zh. Strukt. Khim.*, **35**, No. 6, 176-182 (1994).
6. I. G. Zenkevich, *Abstracts of Papers from the 16th International Symposium on Capillary Chromatography*, Vol. 1, Heidelberg (1994), pp. 41-50.
7. M. S. Vigdergauz, L. V. Semenchenko, V. A. Erzets, and Yu. N. Bogoslovskii, *Qualitative Gas Chromatographic Analysis* [in Russian], Khimiya, Moscow (1978).
8. I. G. Zenkevich, *J. Ecol. Chem.*, **3**, No. 2, 111-119 (1994).
9. A. S. Bureiko, I. G. Zenkevich, and B. V. Ioffe, *Vestn. Sankt-Peterb. Univ., Ser. Fiz. Khim.*, No. 1, 57-61 (1992).
10. *Handbuch der Gas-Chromatographie*, E. Leibnitz and H. G. Struppe (eds.), Geest and Portig, Leipzig (1970).
11. T. A. Misharina, R. V. Golovnya, and I. V. Beletskii, *Izv. Akad. Nauk, Ser. Khim.*, No. 7, 1224-1227 (1993).
12. T. A. Misharina, Sorption of Sulfur- and Oxygen-Containing Compounds in Chromatography: Tendencies and Application for Identification of Volatile Organic Compounds, Author's Abstract of Doctoral Dissertation, Moscow (1995).
13. I. G. Zenkevich, *Zh. Org. Khim.*, **28**, No. 9, 1827-1840 (1992).
14. V. M. Potapov, *Stereochemistry* [in Russian], Khimiya, Moscow (1988).
15. I. G. Zenkevich and B. V. Ioffe, *Interpretation of Mass Spectra of Organic Compounds* [in Russian], Khimiya, Leningrad (1986).
16. I. G. Zenkevich, *Abstracts of Papers from the 15th International Symposium on Capillary Chromatography*, Vol. 1, Heidelberg (1993), pp. 180-196.

Translated by L. Smolina