ELSEVIER

Journal of Chromatography A, 678 (1994) 189–200

JOURNAL OF CHROMATOGRAPHY A

Review

Retrieval of structure information from retention index

C.T. Peng[1]

Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, CA 94143-0446, USA and Department of Chemistry and Nuclear Science Facility, San Jose State University, San Jose, CA 95192-0163, USA

First received 10 January 1994; revised manuscript received 26 April 1994

Abstract

The chromatographic identity of a compound can be determined by four parameters, namely, I, A, Z and (GRF). These are interrelated in a linear regression equation, given in the paper as Eq. 8. The retrieval of structural information from retention data requires the introduction of a new meaning to the Kováts retention index, the use of column difference (ΔI) to characterize functional groups, the redefinition of the role of electronegative oxygen and nitrogen atoms, and the division of retention index (I) into contributions from atoms and from functional groups. The separation of retention index (I) into molecular and interaction contributions is a necessary condition for retention index prediction from structure and also for structure information retrieval from retention data. According to Eq. 8 the retention index is uniquely determined by three parameters, namely A, Z and (GRF). For prediction of retention index, the A value is assigned a value of 100 index units (i.u.), the Z value is obtained directly from the compound, and the (GRF) value is pre-calibrated. In Eq. 10, the m and n values represent the pre-calibrated terms for a quantitative structure–retention index relationship. These terms account for the positive and negative retention contributions from polar and polarizable atom groups. All atom groups that are different from methylene and methyl groups will interact with the stationary phase and contribute to retention. The m and n values for various functional, polar and polarizable atom groups and their column differences (ΔI values) are the results of interactions between the solute and the stationary phase and are structure dependent. The interaction increases with increasing polarities of the solute and the stationary phase. The column difference not only reflects the strength of the interaction, but is also characteristic of the functional and polarizable groups. The retrieval of structural information from retention data is equivalent to obtaining Z and (GRF) values from known I and ΔI values, which is straightforward for monofunctional compounds. For multi-functional compounds, additional data will be needed for retrieval of structural information. These can be obtained from derivatization of the unknown compound, from its chemical reactions with other reagents, from GC–MS analysis and from structure match using internal or external standards. The additional data required will depend upon the complexity of the unknown structure. This approach demonstrates that a system can be devised to utilize GC retention characteristics uniquely for structure elucidation.

[1] Address for correspondence: Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, CA 94143-0446, USA.

0021-9673/94/$07.00
SSDI 0021-9673(94)00431-8

Contents

1. Introduction

Gas chromatography (GC) can separate complex mixtures of closely related components into individual chromatographic peaks by differential adsorption and partition on a chromatographic column but it provides no systematic means for identifying these peaks except by co-chromatography with authentic samples or by mass spectrometric (MS) analysis. The fact that each compound shows a characteristic emergence or retention time corresponding to the structure under given GC conditions, implies that there is structural information in GC retention data. The question is how can a system be devised to extract structural information from retention index. This report will focus on this subject, and the content will be divided into three parts. The first part will review briefly the current status of correlation between retention and structure for predictive purpose to show that not all correlations can be used to retrieve structural information from retention data. The second part will review our work on the prediction of retention index from structure and will point out the conceptual difference between the conventional approach and our approach, which can lead to a system that can be applied conversely. The third part will discuss the retrieval of structural information from retention data and will show a simple procedure for identifying monofunctional compounds. For more complex molecules, additional information from chromatographic retention on different stationary phases, from chemical reactions of the unknown compounds, from GC–MS and from other sources are needed for elucidating unknown structures.

2. Historical background

Gas–liquid chromatography [GLC or may be abbreviated as gas chromatography (GC)] was introduced in 1952 by James and Martin [1]. Its high resolving power and the ease of operation have rapidly gained world-wide acceptance. According to Takács and co-workers, the number of papers published on GC and its applications exceeded 70 000 up to 1983 [2], and there were over 100 important retention index research establishments in 28 countries over the world in 1989 [3]. The Twelfth Collective Index of *Chemical Abstracts* listed over 9000 references under Gas Chromatography in a 5-year span from 1987

to 1991. Numerous reviews, books and conference proceedings on the topic were published [2–6]. These activities are testimony to the importance of GC as a separation and analytical tool.

Aside from various applications, the central interest of GC research has been (i) the study of the influence on retention of carrier gas flow-rate, sample size, temperatures of injector, column and detector, column length, the nature, polarity and film thickness of stationary liquid phase, and the nature of solid support, etc., (ii) the use of retention index systems for reporting of retention data free of influences of experimental parameters for reproducibility and inter-laboratory comparison and (iii) the correlations between retention data and structure and between retention data and physico-chemical properties of the compounds under study. With the introduction of modern column technology and modern GC instrumentation, experimental parameters can be precisely controlled and reproduced and no longer represent a central focus of research interest. The retention data are reported in the literature in terms of equivalent chain length, methylene unit, relative retention with respect to *n*-nonane, retention index, etc. These retention values, while free of the influence of experimental parameters, are difficult to interconvert among one another. The Kováts retention index system, using *n*-alkanes as calibration standards, is the most favored and widely adopted. Other polar calibration standards, such as 1-alkanols [7], 2-alkanones [8], propyl ethers [9], fatty acid methyl esters [10], etc. have been recommended. Their usage is not suitable for prediction of retention index from structure and conversely for retrieval of structural information from retention data. Correlations of retention, structure and physico-chemical properties have been intensively studied. For clarity we will review them briefly separately below:

2.1. Correlation of retention index and physico-chemical properties

Retention and retention index and their correlations with physico-chemical properties have been the subject of many reviews [2–6]. According to Evans and Haken [4], "all the correlations of retention indices and the various physico-chemical properties are of relatively short order (i.e, there are few correlations), or with application being restricted to a particular functional class of functional classes", and "despite much work and many reports it is obvious that no realistic scheme of wide application is available for the precalculation of retention indices". Substantive review articles by Takács and co-workers [2,3] and by Evans and Haken [4] cited over 2200 references and discussed many aspects of retention and all retention index systems. These systems known as generalised retention index, homologous index, unified retention index, standard retention index, invariant retention index, universal retention index, molecular retention index, dispersion and selectivity indexes [2–6] and electric topological index [11], correlate retention with column temperature, boiling points, flow-rate, equivalent molecular mass or physico-chemical quantities, etc.. Evans and Haken [4,6] have given a concise, thorough account of different systems for reporting retention data and the retention index systems in GC. The application of these retention index systems has been limited generally to hydrocarbons on non-polar stationary phases. When retrieving structural information, retention data on at least two columns of different polarities are required to determine the nature of functional groups in the unknown molecule. If the correlations from the above systems are polynomial in nature or only valid for a non-polar column, then these correlations will be unsuitable to use for retrieving structural information from retention data.

Recent studies using molecular descriptors have predicted the retention indexes of 86 olefins and 144 diverse drugs on non-polar columns [12,13]. The molecular descriptors are derived for the molecular structure from a set of parameters consisting of electron density, charge separation, molecular mass, X moment of inertia, molecular connectivity, molecular refractivity, partition coefficient, etc., and are assigned numerical values for a given class of compounds to represent a molecule's properties. This multidescriptor approach can predict the retention in-

dexes of polychlorinated biphenyls [14,15] and nitrated polynuclear hydrocarbons [16]. According to Ong and Hites [17], a linear combination of molecular polarizability, the ionization potential and the square of the dipole moment of the molecule can predict the retention indexes of polycyclic aromatic hydrocarbons, polychlorinated biphenyls, polychlorinated dibenzo-*p*-dioxins and dibenzofurans. These examples show that although the physico-chemical properties of these compounds can be correlated with GC retention data, their relationships cannot be utilized to retrieve structural information from retention data.

2.2. Correlation of retention index and structure

The earliest structure–retention relationship in GC is the correlation of retention and the carbon number of members of a homologous series [18]. In the isothermal mode, the components of a homologous series will emerge logarithmically spaced from the adjacent peaks. Plotting the logarithmic retention time ($\log t_R$) against the carbon number (n) of the homologues gives a linear relationship, as shown below:

$$\log t_R = an + b \quad (I = 100n) \tag{1}$$

where a and b are constants. The equation in the bracket, $I = 100n$, defines the Kováts I values for the retention calibration standards, n-alkanes [19]. The Kováts I value is calculated from the isothermal retention time by the following equation:

$$I = 100i \cdot \frac{\log X_t - \log t_n}{\log t_{n+i} - \log t_n} + 100n \tag{2}$$

where n represents the number of carbon atoms in n-alkanes used as markers; X_t, t_n and t_{n+i} are the adjusted retention times (corrected for the air peak) of the analyte solute, the n-alkane marker with n carbon atoms eluting before and that with $n+i$ carbon atoms eluting after the analyte, respectively; i usually has the value of 1 or 2.

In temperature-programmed GC, the temperature is linearly increased with time, and the components of a homologous series emerge approximately equally spaced from adjacent peaks. Plotting the adjusted retention time (t'_R) with the carbon number (n) gives the following linear regression equation [20]:

$$t'_R = cn + d \tag{3}$$

where c and d are constants. The retention index can be calculated from the temperature-programmed retention time by the following equation:

$$I = 100i \cdot \frac{X - M_n}{M_{n+i} - M_n} + 100n \tag{4}$$

where n is the number of carbon atoms in n-alkanes used as markers; X, M_n and M_{n+i} are, respectively, the retention times of the solute, the n-alkane marker with n carbon atoms eluting before and that with $n+i$ carbon atoms eluting after the analyte, respectively; i usually has the value of 1 or 2.

Both Eqs. 1 and 3 are basic equations relating retention time to carbon number. The four constants a, b, c and d are arbitrary constants with no obvious significance to structure. The intercept b of Eq. 1 has been equated to a thermodynamic quantity [21].

A different approach to correlate retention with structure is adopted in the use of retention increments or the retention indexes of molecular fragments to compose the retention index of a molecule for which the authentic example is not available. Schomburg and Dielmann [22–24] applied retention increments to predict the retention indexes of saturated and unsaturated cyclopropane hydrocarbons by means of the Kováts indexes of isomers or of compounds with the same number of carbon atoms but with cyclopropane ring, double bond, chain branching, etc. in different positions of the molecule. Schomburg [24] also applied this correlation to aliphatic acid methyl esters and showed that different functional groups have characteristic column differences (ΔI values). Cook and Raushel [25] and West and Hall [26] used the same approach to pre-calculate retention indexes of benzene and benzene derivatives. Buchman et

al. [27] identified substituted cyclohexenes formed in tritium labeling using predicted retention index. Dimov and Moskovkina [28] correlated the retention with the structure of benzodiazepines. In the molecular fragment approach, the correlation found for one single class of compounds cannot be extended to other classes. It limits the general application of the retention increments for the molecular fragments. Our studies show that the retention increments contain mixed contributions of interaction and molecular retention.

2.3. A theoretical approach to retention index (I)

Garcia-Raso et al. [29] studied the GC behaviour of alkenes based on molecular orbital calculations, using the total energy, binding energy, energies and coefficients of highest occupied and lowest unoccupied molecular orbitals, etc. to arrive at the conclusion that the retention index (I) can be represented as:

$$I = I_m + I_i \quad (5)$$

where I_m is the molecular contribution and I_i the contribution from solute–stationary phase interaction. It shows that a linear combination of these two retention contributions can be used for structure information retrieval.

Attempts to sub-divide the retention index were made by Takács and co-workers [30–32] who divided the retention contribution into three components, thus:

$$I^{\text{st.ph.}}_{\text{substance}}(T) = I_a + I_b + I_i^{\text{st.ph.}}(T) \quad (6)$$

where I is retention index under isothermal conditions, at column temperature T, I_a atomic index contribution, I_b bond index contribution, and I_i interaction index contribution from a given stationary phase (st.ph.). Without the bond index contribution term I_b, this equation would have been identical to Eq. 5 given above. Souter [33] commented on the impracticality of using the bond index contribution in Eq. 6 for predicting retention index.

Evans and Smith [34,35] divided the retention index into two components, thus:

$$I = I_M + I^* \quad (7)$$

where I_M is the dispersion index, also known as the molecular index, defined as the retention index of a hypothetical n-alkane having the same molecular mass as the solute (M_e). I^* is the selectivity index which reflects the combined effects of molecular shape and functionality and is also given as the carbon number equivalent of ΔM_e. Eq. 7 appears to be identical to Eq. 5 in appearance but the meaning of the terms in these equations differs widely. The concept of effective molecular mass of the solute M_e was first introduced to correlate the relative retention index based on n-nonane (R_{09}) and the molecular mass of the compound (M) by the relation $\Delta M_e = M_e - M$. The correlation achieved by this system is between retention and molecular mass rather than between retention and structure. Retrieving structural information from retention data by the molecular and selectivity index systems, under such circumstances, would be difficult if the molecular mass is not known.

From the above brief survey, one may perceive that these retention index systems provide no direct link of retention data to structure in a manner that allows the process to work conversely to retrieve structural information from retention data.

3. Structure and retention

When the Kováts retention index (I) of members of a homologous series is plotted against the number of atoms in the molecule (Z), a straight line is obtained. We found that his linear correlation holds for homologous series of acids, alcohols, esters, amines, aromatic hydrocarbons, etc. on both non-polar and polar columns [36,37]. The linearity of the plot is expected because each member of the homologous series differs from its nearest neighbors by a methylene group, thus allowing the same retention mechanism to prevail and leading to the observed

linear relationship between the retention index value and the number of atoms (Z) in the molecule. This linear relationship may be represented by the following linear regression equation:

$$I = AZ + (\mathrm{GRF})_Z \quad (8)$$

where A is the regression coefficient and (GRF) the intercept. The $(\mathrm{GRF})_Z$ stands for the group retention factor when the atom number is Z. Eq. 8 is essentially identical to Eq. 5. The AZ term represents the molecular contribution and the (GRF) term the interaction contribution. Eq. 8 is the basis for predicting retention index from structure [36,37] and conversely, for retrieving structural information from retention data.

3.1. The applications of Eq. 8

The four parameters I, A, Z, and (GRF) in Eq. 8 can characterize the chromatographic identity of a compound. According to the Kováts convention, A can be arbitrarily assigned a value of 100 i.u. The three remaining parameters I, Z and (GRF) can be determined under given conditions. Eq. 8 can be used (i) to determine the (GRF) values of functional groups, (ii) to predict the retention index (I) from structure, and (iii) to retrieve structural information from retention data.

3.2. Redefinition of terms

Separation of the retention index into molecular and interaction contributions is essential, in order to ensure the general application of the (GRF) values. Eq. 8 has been successfully applied to predict the retention indexes of compounds of many functional classes on non-polar and polar columns [36–38].

3.2.1. The Kováts index (I).

The Kováts index is unbiased, that is, the index changes when the structure has a different connectivity and is different from methylene and methyl groups. For retrieval of structural information from retention data only unbiased retention index can be used. The Kováts retention index system uses chemically inert n-alkanes as calibration standards [19] and can detect structure features that are different from methylene and methyl groups. The concern that n-alkanes are poorly soluble on polar columns and adsorbed at the liquid–solid interface [40], and that a set of polar compounds can serve as better retention calibration standards for polar compounds on polar columns, may be overly cautious because the adsorption at the liquid–solid interface is difficult to measure, and the poor solubility occurs only when the column is overloaded. From the viewpoint of chemical inertness and general utility, the n-alkanes remain the least problematic retention calibration standards for information retrieval. Polar calibration standards can generate biased retention index, and their use is unsuitable for structure information retrieval.

In element-specific and electron capture detectors which are insensitive to n-alkanes, the use of polar retention calibration standards, such as 1-bromoalkanes [41], 1-nitroalkanes [42], n-alkyl trichloroacetates [43], etc. may be necessary. Conversion of retention data from one system to another is available [20,42].

Conventionally, the Kováts index marks the interpolated position between the time limits set by two adjacent n-alkanes. When used for structural information retrieval, the value of the retention index can also convey a sense of structure. For example, one may view the molecule of benzene as hypothetically formed from n-hexane; it has a retention index of 654 i.u. on a DB-1 column, of which 600 i.u. is the molecular contribution from the 6 carbon atoms, and 54 i.u. is the interaction contribution from the phenyl ring. On DB-Wax column the retention index (I) for benzene is 950 i.u., of which 600 i.u. is the molecular contribution from the 6 carbon atoms, and 350 i.u. is the interaction contribution from the phenyl ring. When predicting retention index from structure, it is necessary to have the structure built up from an n-alkane through a number of change steps. The predicted retention index will be a sum total of the molecu-

lar contribution and the interaction contributions from all the change steps.

3.2.2. The AZ term

The regression coefficient A in the AZ term in Eq. 8 is defined as the retention index increment for atom addition [36,37]. This value is arbitrarily assigned a value of 100 i.u. according to the Kováts convention, meaning that the addition of a carbon atom to the molecule will increase the retention index by 100 i.u. This rule is generally valid for all homologues and all classes of compounds when the methylene group is not influenced by adjacent electronegative groups. The A value will change if the connectivity of the carbon atoms differs from those of methylene and methyl groups, such as chain branching, quaternary carbon atom, etc. The A values are usually less than 100 i.u. when compounds contain adjacent functional groups or highly electronegative groups [36,37,44]. The carbon atoms in the alkyl chain of the phenylalkanes have A values slightly higher than 100 i.u. [45]. The silylated acid amide homologues have the lowest observed A value of 55 i.u. [44].

The parameter Z in the AZ term is the atom number or the number of atoms which include carbon as well as oxygen, nitrogen and other atoms in the molecule [36]. Inclusion of all the atoms in Z is a significant departure from the past convention which uses only the number of carbon atoms for structure correlation. The mass of oxygen and nitrogen atoms contributes to molecular retention, and at the same time, the non-bonding electrons in O and N atoms interact strongly with the stationary phase and contribute to interaction retention. All the functional groups contain O or N atoms or both. If the carbon number (n) is used in Eq. 5 for Z, the regression coefficient A will not change, but the (GRF) value will increase, in which case the (GRF) value will contain mixed interaction and molecular retention components. As such it will vitiate the devised system for retention index prediction from structure and conversely for retrieval of structural information from retention data.

3.2.3. The (GRF) term

The (GRF) term is the group retention factor or the functionality constant of functional groups. The (GRF) values for the functional groups are considerably larger than those for atom groups containing only carbon atoms. Functional groups containing two oxygen atoms or one nitrogen and one oxygen atom, such as carboxyl or acid amide group have larger (GRF) values than functional groups containing one nitrogen atom or one oxygen atom, such as $-NH_2$, $-OH$, $-CHO$ or $-CO$ groups. The non-bonding electrons can interact strongly with the stationary liquid phase. This interaction is characteristic of the functional group. The functional groups on non-polar DB-1 column can be arranged in the order of decreasing (GRF) values as follows [38]:

acid amides > acids > primary alcohols > primary amines > secondary amines = secondary alcohols > aldehydes, ketones > tertiary alcohols > tertiary amines > esters.

The polarity and polarizability of the functional groups can be modified by derivatization. The (GRF) value of the highly polar carboxyl group is about 257 i.u. on a DB-1 column and about 994 i.u. on a DB-Wax column. The former decreases to zero when the carboxyl group is esterified. The ester group has two oxygen atoms which can still interact with the polar stationary phase. This residual polarizability gives the methyl ester a (GRF) value of about 260 i.u. on the polar DB-Wax column.

The (GRF) values for functional groups are usually obtained from a homologous series. When the homologues for a given functional group are unavailable, the retention index increment value (δI) can be used instead. The δI value is the difference in retention index between a compound with the substituent ($I_{subst.}$) and the one without it (I_0) on the same stationary phase, and it should exclude any molecular contribution from atoms [36], thus:

$$\delta I = I_{subst.} - I_0 \quad (9)$$

Both δI and (GRF) are interaction terms, representing the interaction between solute and stationary phase.

3.3. An extension of Eq. 8

The linear regression Eq. 8 for predicting the retention indexes of homologues is not applicable to compounds not members of the homologous series. A general equation for prediction of retention indexes is given below [36–38] :

$$I_p = 100Z + \sum m_i - \sum n_j \quad (10)$$

where I_p is the predicted I value, and Z the atom number, which includes carbon atoms and carbon atom equivalents of oxygen, nitrogen and other atoms in the molecule. The term m_i is the group retention factor of any of the following functions and atom groups: acid, alcohol, aldehyde, amine, ketone, phenol, alicyclic and aromatic ring formation and ring fusion, etc. The presence of these functions and groups in the molecule contribute to a positive interaction term. The term n_j is the group retention factor of any of the following groups: quaternary carbon atom, carbon chain branching or tertiary carbon atom, terminal carbon–carbon double bond,"ortho effect b", fluorine atoms, etc. The presence of these groups in a molecule contribute to a negative interaction term. The m_i and n_j are essentially δI values, obtainable from reference compounds. Significant deviation of the A value from 100 i.u. can seriously affect the accuracy of Eq. 10 for retention index prediction.

4. Retrieval of structural information from retention data

Retention data collected from one column do not contain sufficient information for structure elucidation. For monofunctional compounds, at least two columns of different polarities should be used. Comparison of these retention data can reveal the nature of the functional groups. Additional information are obtained by derivatization and chemical reactions to modify the functional groups. From the chromatographic retention characteristics of these derivatives, one can obtain additional information to deduce the structure of an unknown compound.

4.1. The column difference (ΔI)

The column difference (ΔI) is defined as the difference between two I values of the same compound on two columns of different polarities, thus:

$$\Delta I = I_{\text{more polar}} - I_{\text{less polar}} \quad (11)$$

where $I_{\text{more polar}}$ and $I_{\text{less polar}}$ are retention indexes on more polar and less polar columns, respectively. The column difference (ΔI) is characteristic of the functional group and the column polarity. According to the Kováts convention the molecular contribution on these columns will be the same but the interaction contribution will increase with the polarity of the column. Combining Eq. 8 with Eq. 11 gives the following:

$$\Delta I = (\text{GRF})_{\text{more polar}} - (\text{GRF})_{\text{less polar}} \quad (12)$$

The largest value of ΔI will be between polar DB-Wax and non-polar DB-1 columns. Huber et al. [46] showed that chemical warfare agents, precursors and decomposition products can be identified using a number of stationary phases of low correlation coefficient and that these columns of different retention characteristics can be selected by applying information theory. This indicates that column difference (ΔI) is an important source for structural information. The Rohrschneider and McReynolds constants for characterizing the selectivities of various stationary liquid phases [47–49] are essentially column differences of selected solutes, such as benzene, nitropropane, butanol, pyridine, etc. on polar and non-polar columns. The column difference was first used to define the degree of unsaturation in fatty acids by James [18].

The procedure based on the use of column difference (ΔI) to identify the functionality of an unknown compound from retention data is given as follows: (i) chromatograph the sample on polar and non-polar columns; (ii) compute the retention indexes from the retention times on both columns using Eq. 4; (iii) obtain the column difference by Eq. 11 and compare this value with the values on the list of compiled column differences for functional groups. If the column

difference (ΔI) of the unidentified functional group is found to match that of the primary alcohol group, this information can be confirmed by derivatization; (iv) silylate the unknown compound by preparing trimethylsilylated and *tert.*-butyldimethylsilylated derivatives; (v) chromatograph the silylated derivatives separately on both polar and non-polar columns; (vi) confirm that all polar groups have been masked and that the trimethylsilyl and *tert.*-butyldimethylsilyl derivatives have identical I values on polar and non-polar columns, a condition characteristic of the silylated ether derivatives from alcohols; (vii) obtain the number of atoms (Z) of the unknown alkanol from the I values of these silylated derivatives by dividing by 100. This gives Z and (GRF) values, which concludes the identification.

4.2. The (GRF) and ΔI values and structure

The example given above for retrieving structural information from retention data is based on Eq. 8 which applies only to monofunctional compounds. For compounds containing other polarizable atom groups in addition to functional groups, Eq. 10 is used. This equation contains a number of (GRF) values in summation terms. To determine each of the (GRF) values, additional data will be required.

The magnitudes of the (GRF) and ΔI values are interrelated. Functional groups give large (GRF) and ΔI values. The polarizable atom groups containing only carbon atoms, such as alicyclic and aromatic rings, fused rings, conjugated systems, branched chain, ternary carbon atoms, etc. have small (GRF) and ΔI values. Conjugated systems containing π-electrons can give large ΔI values. A knowledge of the presence of these groups in the molecule can greatly simplify the process of identification. The use of an element-selective detector to detect the presence of nitrogen and halogen atoms and also conjugated bond systems will be extremely useful. The values of (GRF) and ΔI can characterize the polarizable atom groups by the ΔI values from a number of columns of different polarities. Structural atom groups can also be classified into six classes, as listed in the following from groups 1 to 6 in order of approximately increasing ΔI values. These values are based on DB-Wax and DB-1 columns [36,37]:

(1) Carbon atoms with different bond linkages (molecular connectivity), such as double and triple bonded and branch-chained carbon atoms, etc.

(2) Halogen atoms in the molecule.

(3) Aggregates of carbon atoms in single ring formation and polynuclear ring formation.

(4) The configuration of the whole molecule.

(5) An isolated functional group.

(6) A multiplicity of functional groups in the molecule.

This list is incomplete. Only a small portion of ΔI values for the various functional and polarizable atom groups in the above classes are known [36,37,44]. This will be an area for future investigation.

4.3. Retention index and molecular structure

The advantage of correlating retention index and structure is that a set of rules can be formulated from the published data to serve as guidelines, either to predict retention index from structure or to retrieve structural information from retention data [36]. These rules are:

(1) The retention index of a molecule containing Z atoms cannot be less than its base value (i.e., $100Z$) unless the molecule contains fluorine atoms, quaternary carbon atoms and derivatized functional groups in close proximity.

(2) Molecules which contain multiple O and N atoms will have higher I values than molecules that do not.

(3) Molecules which have highly conjugated systems containing N and O atoms tend to give higher I values than those that do not.

(4) Molecules which contain quaternary carbon atoms and functional groups connected to secondary or tertiary carbon atoms have lower I values than those that do not.

(5) Highly substituted molecules tend to yield lower I values than those that do not.

The above rules show that nitrogen and oxygen atoms by virtue of their non-bonding elec-

trons can form hydrogen bond and interact strongly with polar stationary phase by dipole–dipole, dipole–induced dipole, etc. interactions to give high values of interaction retention. Implicit in the observed data is that electron density, molecular planarity and molecular compactness (i.e., spherical shape) can play a very important role in determining the retention of a molecule.

4.4. Derivatization and chemical reactions

Highly polar compounds, such as amino acids, sugars, etc. cannot be analyzed by GC without derivatization. Their polarity can be modified by acylation, methylation, alkylation and silylation. Derivatization can mask the functionality of the polar groups to such an extent that these derivatives behave chromatographically as *n*-alkanes and have practically the same *I* values on polar and non-polar columns [44]. In the case of alcohols, the molecule has only one oxygen atom in the functional group, and the silyl group is highly electronegative so that after silylation the silylated ether has no residual polarity or polarizability and can behave chromatographically as an *n*-alkane. If the silyl group is replaced with an alkyl group, the resultant ether group will have a residual polarizability on polar column to give a small (GRF) value and also a small ΔI value. Since the nature of the chemical reactions and the structure of the reagents are known, the identity of the functional group can be deduced from the retention indexes (I) and the column differences (ΔI) of the derivatives. Different derivatives of the same functional group can be prepared to corroborate the identification. Commonly used reagents for derivatization are: (i) bromine (for probing double bond), (ii) hydrogenation (for probing double bond and reducible groups); (iii) silylating agents (for masking hydroxyl, carboxyl and amino groups): trimethylsilyl- and *tert.*-butyldimethyl silyl-; (iv) acylating agents (for masking hydroxyl and amino groups): acetic anhydride, trifluoroacetic anhydride, pentafluoropropionic anhydride and heptafluorobutyric anhydride; (v) methylating agents (for masking hydroxyl, carboxyl and amino groups): diazomethane, methyl iodide, dimethyl sulfate and alkyl chloroformates [50]; and (vi) Schiff's base formation (for masking amino or keto groups)

This list of reactions and reagents is not exhaustive. The reaction products should be well defined and amenable to chromatography, and their retention characteristics should be understood and can reflect the structure of the unknown compound.

4.5. Structure match using internal or external standards

Another source from which to obtain structural information is by comparing and matching the chromatographic retention characteristics of a known structure with that of an unknown structure under various conditions. Complex molecules such as natural products contain oxygenated ring structures. The oxygen atom can appear as ring oxygen as well as hydroxyl and carboxyl group. The functional groups can be derivatized to eliminate their retention contribution. But the (GRF) values of the atom groups, such as the ring system and the ring oxygen atom would not be affected by derivatization. These (GRF) values will vary when the compound is chromatographed on columns of different polarities. If the changes in (GRF) and ΔI values for the standard and the unknown compound can match each other on various columns, it would be a good indication that the standard and the unknown may have similar structural features.

4.6. Applications

One of the great challenges in analytical and organic chemistry is structure elucidation. UV/IR spectrophotometry can identify structures of chromophores by absorption bands. MS can identify unknown structures by fragment ions and molecular ions. Identification by these measurements is not without ambiguities because UV/IR spectrophotometry can only identify the chromophoric group but not the rest of the molecule, and MS has difficulty in differentiating not only between isomers but also between

diverse compounds that have similar fragmentation patterns [39]. In comparison, high-resolution GC can separate both homologues and isomers. The retention index (I) and column difference (ΔI) can reveal the size of the molecule and the nature of the functional groups. Derivatization and chemical reactions can yield additional structural information about the polarizable atom groups. In structure elucidation, any information is useful. For example, in the study of the mechanism of tritium labeling radioimpurities and radioactive by-products can be successfully identified by the use of quantitative structure–retention index relationship because these products are derived from one known compound.

By convention, GC serves only as a separation tool for the GC–MS technique and not as a source for structure information, but GC can complement MS and contribute to structure elucidation. For example, the search of a mass spectrum library for a structure to match the mass spectrum of an unknown compound can result in several different possible structures. The correct structure is selected on the basis that its predicted retention index and column difference (ΔI) values match the observed values of the unknown chromatographic peak [39]. Retrieving structural information from retention data can be a more efficient way for structural analysis because this method is not limited to analyzing one unknown compound at a time. It can simultaneously analyze many unknown chromatographic peaks in a multi-component unknown mixture, since all the components are similarly processed and chromatographed on various columns, and all the retention data are available for characterization of the unknown peaks. This approach has been successfully applied to identify solvent components in liquid scintillation cocktails [39].

5. Conclusions

With the use of an unbiased retention index system and a quantitative structure–retention index relationship one can obtain structural information from GC retention data. In this system the four parameters I, A, Z and (GRF) given in Eq. 8 can define the chromatographic identity of a compound. Based on this system, it is straightforward to identify an unknown monofunctional compounds from retention data, but to identify an unknown multi-functional compound from retention data alone may depend on the complexity of the molecular structure and the number of atom groups it contains. Each atom group needs a piece of chromatographic information to identify it uniquely. Chromatographic information generated from derivatization and chemical reactions of the unknown compound, from column differences on different stationary phases can identify polar and polarizable atom groups. Together with related information from GC–MS and other sources it can lead to structural identification. MS analysis does not always provide sufficient information for structural assignment, but a combination of analytical data from GC and MS can overcome the shortcoming and elucidate structures of unknown compounds.

Acknowledgement

This publication was made possible by grant number CA 33537 from the National Cancer Institute.

References

[1] A.T. James and A.J.P. Martin, *Biochem. J.*, 50 (1952) 679.
[2] M.V. Budahegyi, E.R. Lombosi, T.S. Lombosi, S.Y. Mészaros, Sz. Nyiredy, G. Tarján, I. Timár and J.M. Takács, *J. Chromatogr.*, 271 (1983) 213.
[3] G. Tarján, Sz. Nyiredy, M. Györ, E.R. Lombosi, T.S. Lombosi, S.Y. Budahegyi, S.Y. Mészáros and J.M. Takács, *J. Chromatogr.*, 472 (1989) 1.
[4] M.B. Evans and J.K. Haken, *J. Chromatogr.*, 472 (1989) 93.
[5] E. Kováts, *Adv. Chromatogr.*, 1 (1965) 229.
[6] J.K. Haken, *Adv. Chromatogr.*, 14 (1967) 367.
[7] A. Gröbler, *J. Chromatogr. Sci.*, 10 (1972) 128.
[8] R.G. Ackman, *J. Chromatogr. Sci.*, 10 (1972) 535.
[9] S.J. Hawkes, *J. Chromatogr. Sci.*, 10 (1972) 536.

[10] F.P. Woodford and C.M. van Gant, *J. Lipid Res.*, 1 (1960) 187.
[11] K. Ośmialowski, J. Halkeiewicz and R. Kaliszan, *J. Chromatogr.*, 361 (1986) 63.
[12] R.H. Rohrbaugh and P.C. Jurs, *Anal. Chem.*, 57 (1985) 2770.
[13] R.H. Rohrbaugh and P.C. Jurs, *Anal. Chem.*, 60 (1988) 2249.
[14] A. Robbat, Jr., G. Xyrafas, and D. Marshall, *Anal. Chem.*, 60 (1988) 982.
[15] M.N. Hasan and P.C. Jurs, *Anal. Chem.*, 62 (1990) 2318.
[16] A. Robbat, Jr., N.P. Corso, P.J. Doherty and D. Marshall, *Anal. Chem.*, 58 (1986) 2072.
[17] V.S. Ong and R.A. Hites, *Anal. Chem.*, 63 (1991) 2829.
[18] A.T. James, *J. Chromatogr.*, 2 (1959) 552.
[19] E. Kováts, *Helv. Chim. Acta*, 206 (1958) 1915.
[20] H. Van den Dool and P.D. Kratz, *J. Chromatogr.*, 11 (1963) 463.
[21] R.V. Golovnya and D.N. Grigoryeva, *Chromatographia*, 17 (1983) 613.
[22] G. Schomburg and G. Dielmann, *Anal. Chem.*, 45 (1973) 1647.
[23] G. Schomburg and G. Dielmann, *J. Chromatogr. Sci.*, 11 (1973) 151.
[24] G. Schomburg, *J. Chromatogr.*, 14 (1964) 157.
[25] L.E. Cook and F.M. Raushel, *J. Chromatogr.*, 65 (1972) 556.
[26] S.D. West and R.C. Hall, *J. Chromatogr. Sci.*, 13 (1975) 5.
[27] O. Buchman, G.-Y. Cao and C.T. Peng, *J. Chromatogr.*, 312 (1984) 75.
[28] N. Dimov and M. Moskovkina, *J. Chromatogr.*, 552 (1991) 59.
[29] A. García-Raso, F. Saura-Calixto and M.A. Raso, *J. Chromatogr.*, 302 (1984) 107.
[30] J. Takács, C. Szita and G. Tarján, *J. Chromatogr.*, 56 (1971) 1.
[31] J. Takács, I. Tálas, I. Bernáth, Gy. Czako and A. Fisher, *J. Chromatogr.*, 67 (1972) 203.
[32] J.M. Takács, *J. Chromatogr. Sci.*, 11 (1973) 210.
[33] P. Souter, *J. Chromatogr. Sci.*, 11 (1974) 418.
[34] M.B. Evans and J.F. Smith, *Nature*, 190 (1961) 905.
[35] M.B. Evans and J.F. Smith, *J. Chromatogr.*, 8 (1962) 303.
[36] C.T. Peng, S.F. Ding, R.L. Hua and Z.C. Yang, *J. Chromatogr.*, 436 (1988) 137.
[37] C.T. Peng, Z.C. Yang and S.F. Ding, *J. Chromatogr.*, 586 (1991) 85.
[38] C.T. Peng, *J. Radioanalyt. Nucl. Chem.*, 160 (1992) 449.
[39] C.T. Peng and S.Q. Liu, in J.E. Noakes, F. Schönhofer and H.A. Polach (Editors), *Liquid Scintillation Spectrometry 1992*, Radiocarbon, Tucson, AZ, 1993, pp. 157–164.
[40] D.F. Fritz, A. Sahil and E. Kováts, *J. Chromatogr.*, 186 (1979) 63.
[41] A. Yasuhara, M. Morita and K. Fuwa, *J. Chromatogr.*, 328 (1985) 35.
[42] R. Aderjan and M. Bogusz, *J. Chromatogr.*, 454 (1988) 345.
[43] T.R. Schwartz, J.D. Petty and E.M. Kaiser, *Anal. Chem.*, 55 (1983) 1839.
[44] C.T. Peng, Z.C. Yang and D. Maltby, *J. Chromatogr.*, 586 (1991) 113.
[45] C.T. Peng, R.L. Hua and D. Maltby, *J. Chromatogr.*, 589 (1992) 231.
[46] J.F.K. Huber, E. Kenndler, G. Reich, W. Hack and J. Wolf, *Anal. Chem.*, 65 (1993) 2903.
[47] L. Rohrschneider, *J. Chromatogr. Sci.*, 11 (1973) 160.
[48] W.O. McReynolds, *J. Chromatogr. Sci.*, 8 (9170) 685.
[49] S.K. Poole, B.R. Kersten, R.M. Pomaville and C.F. Poole, *LC · GC*, 6 (1988) 400.
[50] P. Hušek, J.A. Rijks, P.A. Leclercq and C.A. Cramers, *J. High Resolut. Chromatogr.*, 13 (1990) 633.