

CHROM. 11,189

## THE STRUCTURAL ORIGIN OF CHROMATOGRAPHIC RETENTION DATA

MILAN RANDIĆ

*The Ames Laboratory, USDOE, Iowa State University, Ames, Iowa 50011 (U.S.A.)*

(Received March 3rd, 1978)

---

### SUMMARY

Retention data on alkanes are re-examined from a structural point of view with an emphasis on the molecular connectivity. Rather than adopting empirical parameters for selected structural groupings an assumption that bond contributions depend on the formal carbon valencies in a hydrogen suppressed molecular graph is made. Except for several highly branched systems (which give problems in almost all empirical correlations) this single assumption produces a good linear correlation with the connectivity index which compares favorably with alternative schemes using ten (and more) structural empirical parameters. Furthermore, a close examination of the skeletal forms for the worst cases indicates that these are characterized by a larger number of methyl groups three bonds apart, hence by an additional assumption, that such methyl-methyl fragments cause departures from the correlation (given by the quadratic function of the number of fragments) retention indices for all alkanes considered can be well represented by a linear correlation with the connectivity index. The significance of the work is in demonstration of a dramatic reduction of the structural assumptions required. The approach should be considered as complementary to purely empirical correlations of the retention data with a large number of structural indices, pointing rather to the significance of the structural fragments involved than competing in precision with more flexible alternative schemes.

---

### INTRODUCTION

Regularities in the relative positions of the retention-time data in hydrocarbons have been known for a long time and continue to provide interest among chemists concerned with elucidation of the underlying structural factors involved. Kováts<sup>1</sup> studied among the first the dependence between the structure and gas chromatographic data of organic compounds and summarized his findings in the following three important rules (which we present in somewhat modified form). (1) The retention indices of the higher members of a homologous series increase by addition of each methylene group by a constant; (2) the difference in boiling points among two isomers is proportional to the difference in their retention indices (on a non-polar stationary phase) and (3) the retention indices of non-polar compounds or compounds on non-polar stationary phases remain approximately constant. The first rule allowed

Kováts to suggest a scale for measuring retention indices of an arbitrary (branched) hydrocarbon by fixing the position on the scale for  $C_nH_{2n+2}$  normal paraffins to  $100n$ . Subsequently many investigators extended Kováts approach to other compounds, e.g., alcohols, fatty acids, benzene derivatives, etc. Kováts indices are widely used as a scale for correlating other experimental molecular properties, while the simple regularities and parallelism between chromatographic data and boiling points stimulated search for structural factors involved. In this paper we present a contribution in that direction, and in contrast to the prevailing trend in discussions of the retention data and molecular structure, which attempts to recognize the most relevant molecular fragments we are primarily concerned with the kind of assumptions which may be responsible for the additivity observed. Instead of using a dozen structural fragments to which empirically numerical parameters are assigned we will show that a single assumption on the bond contribution, which is determined by the formal valency of the carbon atoms involved in a molecular graph in which hydrogens are suppressed, accounts for the major part in the additive expression for the retention values. Subsequent examination of the cases which show appreciable departure from the correlation suggests another structural component, the number of methyl-methyl groups three bonds apart. With consideration of only these two assumptions one can derive a correlation between the experimentally observed retention values and purely structural factors, which in quality of the prediction compares well with the alternative empirical curve fitting procedures employing numerous parameters.

Before outlining our scheme we would like to indicate some of the limitations of existing schemes, and indeed of any scheme concerned with correlating the chromatographic data with structural information. Kováts indices are based on an experimentally observed increase in the retention peak with increase of the chain in normal paraffins. The derived scale is then used to determine "positions" of branched isomers. Such a scale when used for discussion of other experimental molecular properties (such as the boiling points) incorporates possible irregularities (even if minor) of the retention data and hence can distort or obscure some effects in the other molecular property. For example, it is known that several highly branched isomers, e.g., 2,2,3-trimethylbutane, 2,3,3-trimethylpentane, etc., show a significant departure from the correlation line between Kováts index (retention data) and the boiling points. Is this due to departure in the linearity for retention data, or does it essentially follow from the thermodynamic data? In fact, even if a point lies close to the line we cannot be sure if the corresponding molecular property is well represented by the additivity, since the two selected features (chromatographic volume and the boiling point) may parallel each other, but otherwise they both may depart from the additivity. Of course, if one of them is additive in a selected additivity scheme, the other should also be additive to the precision dictated by the quality of the correlation. So in this respect correlation of one molecular property with another remains of considerable interest as it permits one to establish equivalent correlation and additivity schemes. However, in view of the above discussion, it should be apparent that if one can devise a structural molecular parameter and use such a parameter separately in two correlations, the objections raised would be cleared. Not long ago we suggested such a structural parameter<sup>2</sup>, called initially branching index, as it was applied to the discussion of the branching characteristics in alkanes. However,

because of the generality of the index, it has been subsequently referred to as the connectivity index<sup>3</sup>. The connectivity index (which will be fully discussed in the next section) has been used in ref. 2 in two separate correlations (a) for the boiling points of alkanes (Fig. 1 in ref. 2) and (b) for the empirical Kováts retention index (Fig. 2 in ref. 2). The scatter of the points in the correlation (which are for similar intervals well approximated by a line) is clearly different illustrating variations in measured and possibly factual relevant data, which in a combined correlation would obscure each other.

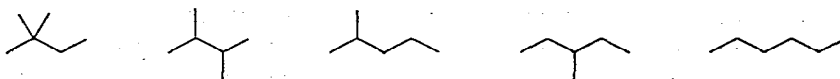
Additivity schemes which relate molecular property to a set of parameters eliminate the above cause of ambiguity. Here one can raise two important questions: (1) What is the minimal number of significant parameters? and (2) how to select among various possibilities of linearly dependent parameters? Statistical analysis is important in the first problem and some pitfalls are possible, in particular in the case of multiple regression unless the number of parameters used is quite small<sup>4</sup>. The problem of selecting structural parameters is more intricate. Apparently different additivity schemes have been proposed for bond properties, to be found later as equivalent<sup>5</sup>. Graph theory is here of importance and the problem has been first investigated by Smolenskii<sup>6</sup> and independently elaborated by Gordon and co-workers<sup>7,8</sup>. Part of the difficulty lies in the fact that no hierarchy of graph invariants which would satisfy strictly the conditions on linear independence have been found<sup>7,8</sup> which then leaves some options in a choice of fragments used as a basis for expansion in an additivity theorem sought. As a consequence, several equally valid and fully equivalent schemes are possible for any additivity scheme which extends the number of fragments used to four atoms or more. One can confine the graph invariants to chain graphs (paths) only, and it appears that for acyclic structures such an approach uniquely characterizes molecular skeleton<sup>9</sup>. Platt<sup>10</sup> has some time ago suggested these path numbers as useful quantities in discussion of properties of isomers, while Altenburg<sup>11</sup> found them useful in construction of polynomials which give an average squared molecular diameter of alkanes. However, extension of the concept to general polycyclic structures, when the number of possible paths increases enormously, remains to be investigated.

Finally, one should be constantly aware of the limitations of correlations among the structural data with molecular property if the latter depends on additional experimental condition. In the case of the chromatographic retention volumes some dependence on the nature of the supporting phase and on the temperature is well documented. Hence some variations, besides a constant shift in the magnitudes (major part of the temperature dependence) which is not troublesome, are clearly beyond any additivity scheme, and indicate the respective environmental influences. For example the specific retention values for 2-methylheptane and 3-methylheptane on  $n$ -C<sub>28</sub>H<sub>30</sub> and  $n$ -C<sub>32</sub>H<sub>34</sub> are reported<sup>12</sup> as 106.5 and 116.6, and 108.9 and 108.6, respectively. Hence, it appears that within 10 retention units the two supporting materials can influence the relative positions of the retention peaks. With this in mind, any scheme which approaches the above limit should be considered practical.

## THEORETICAL

*Connectivity index*

The connectivity index which we introduced for the discussion of variations among isomeric alkanes essentially incorporates in its definition, in a simple manner, some information on the immediate environment for the atoms making the bond. Consider the five isomers of hexane:



These are ordered in an apparent sequence that parallels our intuitive notion of the degree of branching of the corresponding molecular skeletons, although we evaded resolving such ambiguities as whether 2-methylpentane should be considered more branched than 3-methylpentane, or whether at all such a question is legitimate<sup>13</sup>. The particular above ordering is derived from sequencing the unique adjacency matrices that one can derive for the structures considered<sup>2,14,15</sup>. Having an ordered sequence of structures accomplishes the first stage in a search for regularities in experimentally available data. Graph theory as such can provide the basis for alternative orderings, depending which structural feature is selected for considerations, graph theory does not produce additional or alternative data. To go from an ordering to a correlation one needs a parameter (or parameters) which can be assigned to an individual structure. In searching for such parametrization one tries to preserve the ordering already established, if this is found to parallel the ordering of molecular property of interest. We are here concerned with the retention data and are searching for a parametrization which will assign to individual structures numerical values that follow the sequence of experimentally found Kováts indices. It is only natural to start investigating an additive scheme by first examining bond additivities. We classify bonds in various  $(m, n)$  bond types. The ordering of structures (isomers of hexane) imposes a set of inequalities which have to be satisfied if the ordering is to be preserved<sup>2</sup>. For instance, in order that 2-methylpentane precedes 3-methylpentane we have the inequality:

$$(1,2) + 2(1,3) + 2(2,2) + (2,3) < 2(1,2) + (1,3) + (2,2) + 2(2,3)$$

which reduces when equal contributions are eliminated to

$$(1,3) + (2,2) < (1,2) + (2,3)$$

A sufficient number of such inequalities can be generated which guides us in assignment of bond parameters to individual  $(m, n)$  bond type. As pointed out in the ref. 2 a simple assignment of a value  $m^{-\frac{1}{2}}n^{-\frac{1}{2}}$ , i.e., the product of reciprocal square roots of carbon formal valencies, already produces the required effect. Although this may appear as an *ad hoc* selection one may recall that the same magnitudes appear as the coefficients of s orbital participation in construction of atomic hybrid orbitals, hence the suggestion is not fully without basis. Hybrids have been found useful in

calculation of bond overlaps which in turn correlate well with selected bond features<sup>16,17</sup>. Arbitrary character of the connectivity index however cannot be denied, nor should it be concealed. Connectivity index is not an observable quantity and is in essence as arbitrary as is the well known Coulson bond order<sup>18</sup> which is a similarly non-observable quantity. Such quantities serve as a means of comparing different properties in different molecules, or properties attributed to molecular fragments and should be justified by their subsequent use in application. The past application of the Coulson's bond order is generally well known, but similarly the use of the connectivity index has been well demonstrated<sup>19</sup>. We refer here specifically to the examples involving hydrocarbons, and families of hetero substituted derivatives involving a same (or similar) structural and functional group in which case a single connectivity index suffices to account for the major trends in the corresponding molecular additivity. In a number of other applications the authors extend the simple scheme to several indices of different order (*i.e.*, including next distant neighbors). Although such extensions are legitimate, the question of multiple regression used needs a very detailed examination in order to appreciate the extent of the improvement due to a simple increase of the number of parameters and the improvement due to importance of particular extension of the basis.

TABLE I

LIST OF BOND TYPES AND THEIR RESPECTIVE CONTRIBUTIONS TO THE MOLECULAR CONNECTIVITY INDEX

<i>Bond type</i>	<i>Contribution</i>
(1,1)	1.0000
(1,2)	0.7071
(1,3)	0.5774
(1,4)	0.5000
(2,2)	0.5000
(2,3)	0.4083
(2,4)	0.3536
(3,3)	0.3333
(3,4)	0.2887
(4,4)	0.2500

Table I summarizes the information bond types and their contributions to the overall molecular connectivity index. With an addition of a methylene group in normal paraffins the connectivity index increases by a constant 0.5000 which makes comparison of the Kováts scale and the connectivity scale very simple, the two differ by a factor of 200. Since the number of bond types is limited it is apparent that as the size of a compound increases a number of isomers will have an identical bond type composition. This happens for example for 3-methylheptane and 4-methylheptane, or 2-methyl-3-ethylpentane and 3,4-dimethylhexane. Such coincidences are not necessarily a sign of limitation of the scheme, since there is no *a priori* reason why selected molecular properties for some isomers could not be the same or almost identical. Fig. 1 shows the correlation between the connectivity index and the experimentally available retention indices<sup>20</sup>.

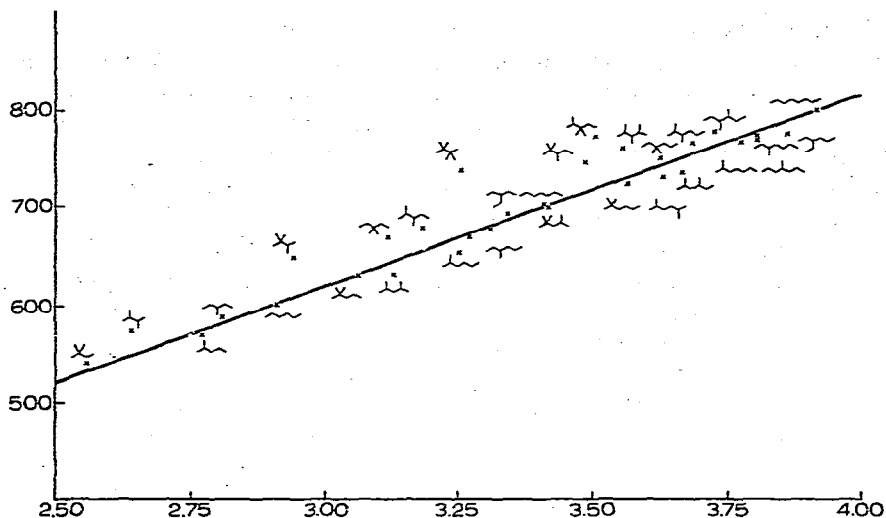


Fig. 1. Correlation between the experimentally reported retention indices (ref. 17) and the molecular connectivity indices for hexane and heptane isomers.

#### Analysis of the correlation

Over a large interval of the connectivity values, the results of Fig. 1 can be well represented by a linear dependence. The line drawn on Fig. 1 has been selected so that it passes through all points belonging to the normal paraffins, rather than representing a least square adjusted line. Such an approach has additional advantages: it divides all paraffins of Fig. 1 into two classes, those whose experimental retention indices are lower than the values expected from interpolation based on linear systems, and those whose retention indices are larger than expected on the same grounds. It is interesting to see that the two classes show some apparent structural differences. Methyl and *m,m*-dimethyl isomers are generally close to the line. Isomers with several methyl substituted groups show generally the greatest departures. In Table II the worst cases are listed showing the amount of the departure from the line. It should

TABLE II

COMPOUNDS WHICH SHOW THE WORST AGREEMENT (MORE THAN 20 RETENTION UNITS) IN THE CORRELATION BETWEEN THE CHROMATOGRAPHIC RETENTION VALUES AND THE MOLECULAR CONNECTIVITY INDEX

Number of methyl-methyl fragments separated by three bonds and the departure of the corrected retention values.

Compound	Diff.	(1,1) <sub>3</sub>	Diff.' (corrected)
2,2,3,3-Tetramethylbutane	75	9	20
2,3,3-Trimethylpentane	55	6	20
2,2,3-Trimethylbutane	40	6	10
2,2,3-Trimethylpentane	30	4	15
2,3,4-Trimethylpentane	30	4	15
2,2-Dimethylbutane	25	4	10
3,3-Dimethylpentane	25	4	10
2,3-Dimethylpentane	20	3	15

be observed that these same isomers give the largest deviations also in number of other additivity schemes and several such results for other schemes are included in Table III for the purpose of comparison.

The question is now to investigate if definite and clear structural features can be detected in the compounds showing the largest deviations in order to see if such deviations have some simple structural basis. A close review of the compounds in Table II shows in all of them a prominent number of methyl groups. However, there are also some compounds having several methyl groups which show a satisfactory agreement with the correlation line. This is the case with 3-methylpentane, 2,2-dimethylpentane and in particular 2,2,4-trimethylpentane. So the question is to recognize differences among the isomers that depart from the line and those, seemingly similar compounds which however show no appreciable departure from the correlation. After a closer look it becomes clear that the two groups, both having one or more methyl groups, differ in the number of fragments having methyls at the end. In particular the fragments with methyl-methyl groups separated by three CC bonds is large for isomers above the correlation line, and generally increase with the departure of the isomer from the line. Isomers which are below the correlation line have no such methyl-methyl fragments or at most one. From this observation, and the magnitudes involved it is simple to derive a substantial improvement in the correlation by "correcting" the retention data, the amount of correction being a function of the number of methyl-methyl fragments separated by three CC bonds. Moreover, examination of the numerical magnitudes for individual departures shows that simple quadratic correction will reduce the departure of the worst cases below 20 retention units. This limit in the precision of the correlation is favorably comparable with several alternative additivity schemes using ten and more structural parameters. Fig. 2 shows the redrawn correlation with the corrected values for the retention indices. The retention indices (RI) are given by the equation:

$$RI = 200 (\chi - 1.4142) + 100 + (T_3)^2$$

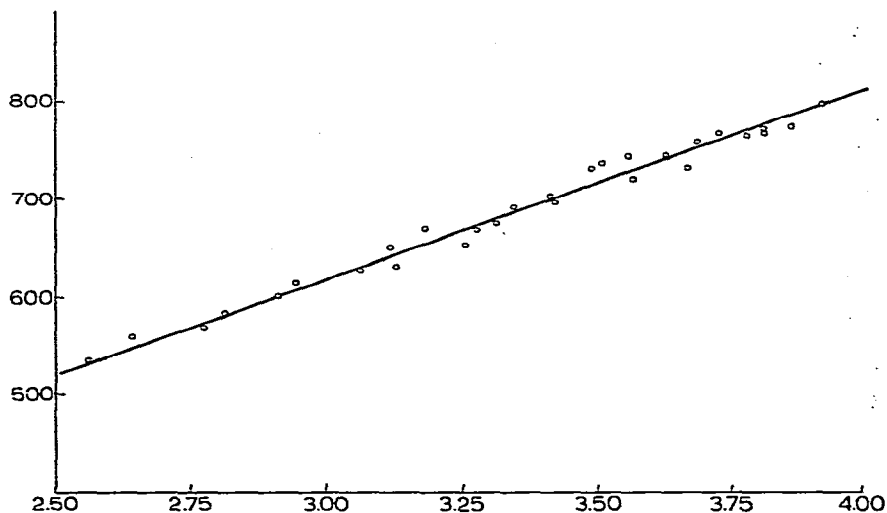


Fig. 2. Correlation of Fig. 1 corrected for excessive departures of the highly branched systems having a number of methyl-methyl groups separated by three CC bonds.

TABLE III

COMPOUNDS AND THE DEGREE OF DISAGREEMENT IN PREDICTION OF THE RETENTION VALUES FOR SELECTED ADDITIVITY SCHEME

Method	Compound	Degree of disagreement	Method	Compound	Degree of disagreement
Takács <sup>21</sup>	2,2-Dimethylbutane	-12	Spivakovskii <i>et al.</i> <sup>21</sup>	2-Methylpentane	-7
	2,2-Dimethylpentane	8		<i>n</i> -Hexane	-10
	3-Ethylpentane	-12		2,2-Dimethylpentane	-8
	2,2,4-Trimethylpentane	-12		2,4-Dimethylpentane	-7
	2,2-Dimethylhexane	-12		2,3-Dimethylpentane	8
	2-Methyl-3-ethylpentane	19		3-Ethylpentane	8
	2,2,4-Trimethylhexane	12		2,2-Dimethylhexane	-8
	2,2,3,4-Tetramethylpentane	-43		3,4-Dimethylhexane	10
	3,3-Dimethylheptane	-21		2,2,4,4-Tetramethylpentane	26
				2,2,5-Trimethylhexane	-11
				2,2,3-Trimethylhexane	-8
				2,2-Dimethyl-3-ethylpentane	-10
				2,4-Dimethyl-3-ethylpentane	-9
				2-Methyl-3-ethylhexane	-11
				3-Methyl-octane	10
		<i>n</i> -Nonane	9		
Castello and co-workers <sup>25,26</sup>	3-Ethylpentane	7			
	2,2,4-Trimethylpentane	9			
	2,3,4-Trimethylpentane	-8			
	2,2,4,4-Tetramethylpentane	42			
	2,2,5-Trimethylhexane	-8			
	2,3,3,4-Tetramethylpentane	15			
	3-Ethylheptane	10			
	2,3-Dimethyl-3-ethylpentane	10			



where RI is the retention index,  $\chi$  is the molecular connectivity index and  $T_3$  is the number of terminal paths of length three, *i.e.*, the number of paths between two methyl groups three bonds apart. The factor of 200 converts the Kováts retention scale to that of the connectivity values, while 1.4142 and 100 are the connectivity index and the retention (Kováts) value for propane.  $T_3$  in fact measures overcrowdedness in a molecule. It appears in the above expression without a numerical factor, but generally one could introduce such a factor by using the data on selected molecules for a least square fit. We have here been primarily interested in indicating how few assumptions may suffice to produce a rather impressive correlation, rather than searching for the best fitting of the data.

## DISCUSSION

The additivity schemes for physico-chemical properties have attracted considerable attention among chemists. A number of publications deals specifically with the chromatographic retention data and propose additivity schemes using different structural parameters for their derivation. We will briefly survey a few representative papers, more extensive literature can be traced through references cited therein (see in particular the list of references in ref. 21). Essentially, authors identify certain molecular fragments, or even bonds, in a specific environment and attribute to these numerical parametric values. Thus Takács<sup>21</sup> discriminates among various CC bonds according to their order in the structure, having for instance in *n*-octane bond types  ${}^0_0C_{21}$ ,  ${}^1_1C_{22}$ , and  ${}^2_2C_{22}$ . To these bond types one attributes specific bond increments, the net retention volume can then be calculated from the knowledge of the molecular structure of the substance. A critical application of the Takács's scheme has been subsequently performed for saturated and unsaturated hydrocarbons<sup>22</sup> (retention indices on squalene at 50° of Rijks and Cramers<sup>23</sup> were adopted). Table III summarizes the information on the compounds which show the worst agreement (*i.e.*, above 10%). It is not so surprising to see that for a number of highly branched species the predictions are less satisfactory. A point that, however, appears not to have been raised is that in a number of other highly branched systems agreement is very good, *e.g.*, 2,2,3,3-tetramethylpentane or 2,2-dimethyl-3-ethylpentane. If one wants to improve the scheme, or understand its limitations it would require some explanation why some seemingly similarly branched systems behave so differently. Vanheertum<sup>22</sup> wonders if the CH contributions, all represented here by a single parameter, should not have been differentiated, and on the other hand, have not some differences among CC bond types been exaggerated? If we compare our numerical values for bond connectivity indices of various bond types, a factor of 2 is not uncommon for selected bond types. So in this respect we would not raise objections. The problem with hydrogen bonds is somewhat more involved. Clearly representing all hydrogens by a single term is a limitation. Among the group of isomers it only represents an additive constant and cannot be responsible for isomeric variations. From our point of view such a parameter is redundant, since the presence of hydrogens has been taken into account via the differentiation of CC bond type, where the formal valency in a hydrogen suppressed graph signifies primary, secondary and tertiary carbons. So it appears to us that the dilemma whether an improvement of the reliability should be based on an extension of the number of bond contributions (Vanheertum<sup>22</sup>) or

simplifications (Souter<sup>24</sup>) will be resolved once we better understand variations of the scheme. In that respect the discussion of the variations shown in Fig. 1 based on the connectivity index and the improvement which results once methyl-methyl fragments three bonds apart are recognized as the major source for the departure in the correlation, may provide a useful guideline. We may mention that Takács<sup>21</sup> made a distinction among some bond types that we do not differentiate, such as for example  ${}^4_2C_{22}$  and  ${}^3_2C_{22}$  which in our case are the same (2,2) bond type.

Castello and co-workers<sup>25,26</sup> considered several physical properties of branched-chain paraffins and calculated the additivity of the retention indices employing a dozen characteristic structural groupings, essentially differentiating various methyl, methylene and ethyl groups. They selected isomers with the relevant groups in their structure and solved the accompanying system of equations (involving 8-11 variables at a time). The agreement between the calculated and the observed retention indices is generally satisfactory (the comparison should be concentrated on nonane isomers as they have not been employed in the calculation of the parameters), except for 2,2,4,4-tetramethylpentane, which is not unexpected, as pointed out by the authors<sup>25,26</sup> due to the fact that this compound contains the first example of a secondary carbon atom between two quaternary groups and the index contribution for this arrangement was not previously determined. Table III lists the worst cases of agreement for calculated retention indices according to Castello and co-workers<sup>25,26</sup>. There are some interesting salient features in the "system" approach of Castello and co-workers, which seem worth mentioning. In their system method several alternative selections of standard compounds have been considered and it was found that the results may be quite sensitive to the set of the parameters adopted (which depends on the selection of the compounds used for their evaluation). However, some of the structural groups appear remarkably insensitive to the selection process: *e.g.*, the role of an external tertiary carbon atom, two quaternary carbon atoms in  $\alpha$ -position, and tertiary and quaternary carbon atoms in  $\alpha$ -position. On the other hand derived parameters for some other selected groupings, such as for example internal tertiary carbon atom, or two ethyls on the same carbon atom show wide discrepancies. In the first case mentioned the values lie in the range from -27 to -71 retention units and from +18 to -16 retention units, respectively. This is definitely a sign that other important factors have not been properly differentiated. The problem of choosing the most important molecular groupings is quite complicated. The above approach of Castello perhaps indicates how to proceed. The difficulty lies in the possibility that some or several structural components are not fully independent. So the large variation in the parameters found by Castello when different "system" equations are used does not mean that those parameters are irrelevant to the problem, more likely they are interrelated. For example values of -27 for internal tertiary carbon correspond to values of +19 for two tertiary carbon atoms, while the corresponding values of -61 and -63 are accompanied by the values of +36 and +37 for the parameters of two tertiary carbons. Other similar regularities can be found on a close examination of the parameters. The selection of the structural groups for the basis of the calculation is left to the intuition and experience of researchers. A more systematic scheme to the question of the selection of the fragments has been recently reported by Spivakovskii *et al.*<sup>27</sup> who discriminate various "forms of bonds" which in turn represent the structural group formed by two bonded atoms and by atoms in its immediate

surroundings. The significance of "bond forms" or "bond types" has been recognized probably by many in the history of structural physical chemistry. The success of the concept of hybrid orbitals of Pauling<sup>28</sup> can be attributed to a proper acknowledgement of the nearest atomic environment. About thirty years ago Hartmann<sup>29</sup> used the concept of bond types in discussion of conjugated hydrocarbons. More recently Dubois and co-workers<sup>30,31</sup> considered the problem more fully and developed a formal treatment of atomic environment which takes into account the number of neighbors a certain distance away from a considered atom. One of the important outcomes of Dubois' systematic approach to molecular structure is recognition of the use of codes (given by a sequence of numbers, as a rule integers) for characterization of a structure. Codes may characterize a structure only fragmentarily, but regularities in isocodal systems (*i.e.*, those having the same codes) when codes signify the same number of primary, secondary, tertiary and quaternary carbon atoms (*i.e.*, a 4-digit code) have been already recognized<sup>32</sup>. In fact the Dubois scheme has been applied to the study of the chromatographic data on paraffins and some derivatives<sup>33-35</sup>. In Table III we summarized the results of Spivakovskii *et al.*<sup>27</sup> for the cases which show the worst agreement (departure in the calculated additivity of the correlation index by 10% or more). Authors used 10 structural groups which involve 2-8 carbon atoms. There are no excessive departures, except 2,2,4,4-tetramethylpentane, which, as we have already seen, obviously introduces difficulties which are not present in less methyl substituted derivatives. Though, again one notices that even more crowded 2,2,3,3-tetramethylpentane gives a fair agreement with the experimental value. One can also observe that the group parametrization of Spivakovskii *et al.*<sup>27</sup> does not preserve the simple linearity of the retention index scale for normal paraffins: The difference between the calculated and the experimental retention indices for *n*-hexane, *n*-heptane, *n*-octane and *n*-nonane are -9.9, -3.5, +2.9 and +9.2, respectively. The departure is not large though and within the limits of the accuracy of the method.

Our own approach is also based on the concept of bond types, however, where we differ importantly from all other investigators is that the differences among bond types are not taken as a basis for parametrization based on an empirical data fitting procedure. Rather, to each bond type an index simply derived from the characterization of the bond is suggested. While in alkanes we have 10 bond types (*m, n*) with *m, n* taking values between 1 and 4 (in fact the bond type (1, 1) can be eliminated from the list as it represents an isolated CC bond, the type which appears only in ethane), the assigned indices are prescribed by a single assumption. The parameters we use are not independent as is the case with other schemes and one should not look at our scheme as a procedure in which numerous parameters are used and somehow disguised. Rather the attitude should be taken: are the numerous parameters that other schemes employ quite independent? or to put it more adequately: are the major contributions of all the distinctive parameters perhaps not correlated? In this respect our approach stands alone as a fundamentally different method. It may appear astonishing that here, where a dozen parameters still leave some room for improvement another scheme based on one or two simple structural assumptions almost competes in precision with more conventional approaches. This, however, should not be so surprising and in fact Wiener<sup>36-38</sup> demonstrated that judiciously selected structural parameters can lead to impressive predictions of selected physico-chemical properties with as few as two selected structural features for encoding of

the molecular structure. The significance of the results of Wiener has been immediately recognized by Platt<sup>39,40</sup> who suggested additional structural numbers that may appear of potential interest in the future. Although some useful results (independently) followed where such structural parameters provided a basis for discussion of molecular properties\* the important message of Wiener has, as it appears, not received proper reception. A possible reason for this may be a lack of appreciation for the importance of an interpretation of the relative magnitudes involved. It is obvious that if more parameters are used a better precision is to be expected. Here graph theoretical approaches differ from empirical curve fitting schemes, the two involve different priorities: the former aims at a clarification of the relative magnitudes involved, the latter aims at a practicality of the predictive scheme, *i.e.*, the accuracy. The schemes of Takács<sup>21</sup>, Castello and co-workers<sup>25,26</sup>, Spivakovskii *et al.*<sup>27</sup> and numerous others not explicitly mentioned here, may be improved by some reconsideration of the parametrization and the fragments used, however, such improvements although very important will not shed light on the derived relative magnitudes. In contrast, the graph theoretical schemes may not satisfy the high accuracy standard necessary for a reliable prediction in identification of the retention peaks. However, the numerical magnitudes used will be related in a more straightforward manner to the pertinent structural factors.

Wiener<sup>36-38</sup>, in his discussion of selected thermodynamic properties of hydrocarbons, found it sufficient to use only two structural parameters  $w$  and  $p$  (besides  $n$ , which determines the number of carbon atoms in a molecule); one determines the number of all paths between all pairs of atoms ( $w$ ), the other represents the number of pairs of carbon atoms separated by three carbon-carbon bonds.

The Wiener parameter might appear as mysterious, but again it can be put in a parallelism with our methyl-methyl three CC bond separated fragments, although they correspond to distinctive structural features. Table IV enumerates all three paths for molecular fragments for hexane and heptane isomers which are differentiated by the formal valencies of their terminal atoms. The  $(1,1)_3$  and  $(1,2)_3$  symbols designate molecular fragments three CC bond lengths with the terminal carbons being primary, and primary and secondary, respectively. The first entry is the structural parameter found essential for correcting the connectivity index in the correlation with the chromatographic retention indices in this work (Figs. 1, 2), while the sum of all entries in a row gives the parameter  $p$  of Wiener. A close examination of these, and similar structural parameters may well lead to a better understanding of the disagreements and agreements in available additivity schemes. It is, of course, outside the scope of the present work to discuss limitations or suggest improvements in other available schemes. Our major emphasis was not to present a scheme to replace the existing models, but to draw the attention of interested scholars to advantages of graph theoretical methods and their potential in considering bond additivity schemes. We would like to emphasize a complementary rather than competitive nature of our scheme based on the concept of connectivity index. It is also possible, if not even likely, that additional contributions will be required in order to improve the quality of such additivity schemes. The concept of bond connectivity index does not go

---

\* *E.g.*, the work of Altenburg<sup>11</sup> on calculation of the radius of branched molecules where polynomials were constructed the coefficients of which are the path numbers that also Platt<sup>10</sup> considers.

TABLE IV

ENUMERATION OF FRAGMENTS INVOLVING THREE CC BONDS BROKEN DOWN INTO COMPONENTS  $(m,n)_3$ , WHERE  $m,n$  REPRESENTS THE FORMAL VALENCIES OF THE TERMINAL ATOMS OF THE FRAGMENTS IN A HYDROGEN SUPRESSED MOLECULAR GRAPH

<i>Compound</i>	<i>Composition (three CC bond paths only)</i>	<i>No. of methyl-methyl fragments*</i>	<i>Total No. of paths length three**</i>
2,2-Dimethylbutane	3 (1,1) <sub>3</sub>	3	3
2,3-Dimethylbutane	4 (1,1) <sub>3</sub>	4	4
2-Methylpentane	2 (1,2) <sub>3</sub> ; (1,3) <sub>3</sub>	0	3
3-Methylpentane	2 (1,1) <sub>3</sub> ; 2 (1,2) <sub>3</sub>	2	4
<i>n</i> -Hexane	2 (1,2) <sub>3</sub> ; 2 (2,2) <sub>3</sub>	0	4
2,2,3-Trimethylbutane	6 (1,1) <sub>3</sub>	6	6
2,2-Dimethylpentane	3 (1,2) <sub>3</sub> ; (1,4) <sub>3</sub>	0	4
3,3-Dimethylpentane	4 (1,1) <sub>3</sub> ; 2 (1,2) <sub>3</sub>	4	6
2,4-Dimethylpentane	4 (1,3) <sub>3</sub>	0	4
2,3-Dimethylpentane	3 (1,1) <sub>3</sub> ; 2 (1,2) <sub>3</sub> ; (1,3) <sub>3</sub>	3	6
2-Methylhexane	3 (1,2) <sub>3</sub> ; (2,3) <sub>3</sub>	0	4
3-Methylhexane	(1,1) <sub>3</sub> ; 2 (1,2) <sub>3</sub> ; (1,3) <sub>3</sub>	1	4
3-Ethylpentane	6 (1,2) <sub>3</sub>	0	6
<i>n</i> -Heptane	2 (1,2) <sub>3</sub> ; 2 (2,2) <sub>3</sub>	0	4

\* Separated by three CC bonds.

\*\* Used in Wiener's analysis<sup>36-38</sup> as the parameter  $p$ .

beyond the nearest neighbor acknowledgement. That this is not always enough is indicated by the need for enumeration of fragments (1,1)<sub>3</sub>, which are just single entries which characterize the connectivity beyond the nearest neighbor limit. A close investigation of other such contributing terms may prove beneficial. A rather systematic approach to a search for regularities in available experimental and theoretical data using the concept of atom and bond codes, rather than atom and bond indices, has recently been initiated, and shows much promise<sup>39,40</sup>. This kind of development offers some optimism for molecular additivity schemes, which may reach not only a satisfactory precision, but at the same time throw enough light on the relative numerical magnitudes employed.

#### ACKNOWLEDGEMENT

This work was supported in part by the U.S. Department of Energy, Division of Basic Energy Sciences.

#### NOTE ADDED IN PROOF

Recently Kaliszan and co-workers<sup>41-43</sup> also examined the use of connectivity indices for correlations with retention indices. In another study Michotte and Massart discussed the limitations when chemically very different groups (alcohols, esters, ethers) were treated. As was to be expected, different functional groups introduced distinctive additive components.

## REFERENCES

- 1 E. Kováts, *Z. Anal. Chem.*, 181 (1961) 351.
- 2 M. Randić, *J. Amer. Chem. Soc.*, 97 (1975) 6609.
- 3 L. B. Kier, L. H. Hall, W. J. Murray and M. Randić, *J. Pharm. Sci.*, 64 (1975) 1971.
- 4 J. G. Topliss, *J. Med. Chem.*, 15 (1972) 1066.
- 5 D. H. Rouvray, *RIC Rev.*, 4 (1971) 173, *Amer. Sci.*, 61 (1973) 729.
- 6 E. A. Smolenskii, *Zh. Fiz. Khim.*, 38 (1964) 1288.
- 7 M. Gordon and J. W. Kennedy, *J. Chem. Soc., Farad. Trans. II*, 69 (1973) 484.
- 8 M. Gordon and W. B. Temple, *J. Chem. Soc., Farad. Trans. II*, 69 (1972) 282.
- 9 M. Randić, in preparation (preliminary report available).
- 10 J. R. Platt, *J. Phys. Chem.*, 56 (1952) 328.
- 11 K. Altenburg, *Kolloid-Z.*, 178 (1961) 112.
- 12 E. C. Pease and S. Thorburn, *J. Chromatogr.*, 30 (1967) 344.
- 13 I. Gutman and M. Randić, *Chem. Phys. Lett.*, 47 (1977) 15.
- 14 M. Randić, *J. Chem. Phys.*, 60 (1974) 3920.
- 15 M. Randić, *J. Chem. Inf. Comp. Sci.*, 17 (1977) 171.
- 16 M. Randić, *Int. J. Quant. Chem.*, 8 (1974) 643.
- 17 M. Randić and Z. B. Maksić, *Chem. Rev.*, 72 (1972) 43.
- 18 C. A. Coulson, *Proc. Roy. Soc. (London) A*, 169 (1939) 413.
- 19 L. B. Kier and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976.
- 20 D. A. Tourres, *J. Chromatogr.*, 30 (1967) 357.
- 21 J. M. Takács, *J. Chromatogr. Sci.*, 11 (1973) 210.
- 22 R. Vanheertum, *J. Chromatogr. Sci.*, 13 (1975) 150.
- 23 J. A. Rijks and C. A. Cramers, *Chromatographia*, 7 (1974) 99.
- 24 P. Souter, *J. Chromatogr. Sci.*, 12 (1974) 418.
- 25 G. Castello, M. Lunardelli and M. Berg, *J. Chromatogr.*, 76 (1973) 31.
- 26 G. Castello and G. D'Amato, *J. Chromatogr.*, 107 (1975) 1.
- 27 G. I. Spivakovskii, A. I. Tishchenko, I. I. Zaslavskii and N. S. Wulfson, *J. Chromatogr.*, 144 (1977) 1.
- 28 L. Pauling, *The Nature of the Chemical Bond*, Cornell University Press, Ithaca, N.Y., 1960.
- 29 H. Hartmann, *Z. Naturforsch. A*, 2 (1947) 259.
- 30 J. E. Dubois, D. Laurent and H. Veillard, *C. R. Acad. Sci.*, 263 (1966) 764, 1245.
- 31 J. E. Dubois and D. Laurent, *Bull. Soc. Chim.*, (1969) 2449.
- 32 A. C. W. van Kemenade and H. Groenendijk, *Chromatographia*, 2 (1969) 316.
- 33 M. Chastrette, G. Lenfant, and J. E. Dubois, *C. R. Acad. Sci.*, 265 (1967) 602.
- 34 G. Lenfant, M. Chastrette and J. E. Dubois, *J. Chromatogr. Sci.*, 9 (1971) 220.
- 35 M. Chastrette, P. Couillault, G. Casteignau and M. Mazet, *Bull. Soc. Chim. Fr.*, (1973) 1605.
- 36 H. Wiener, *J. Amer. Chem. Soc.*, 69 (1947) 17, 2636.
- 37 H. Wiener, *J. Phys. Colloid Chem.*, 52 (1948) 425.
- 38 H. Wiener, *J. Phys. Chem.*, 52 (1948) 1082.
- 39 M. Randić, *Chem. Phys. Lett.*, in press.
- 40 M. Randić, *J. Chem. Soc., Farad. Trans. II*, submitted for publication.
- 41 R. Kaliszan, *Chromatographia*, 10 (1977) 59.
- 42 R. Kaliszan and H. Foks, *Chromatographia*, 10 (1977) 346.
- 43 R. Kaliszan and H. Lamparczyk, *J. Chromatogr. Sci.*, 16 (1978) 246.