

Composition as a method for data reduction: application to carbon-13 NMR chemical shifts*

Milan Randić¹ and Nenad Trinajstić²

¹ Department of Mathematics and Computer Science, Drake University, Des Moines, IA 50311, and Ames Laboratory-DOE**, Iowa State University, Ames, IA 50011, USA

² The Rugjer Bošković Institute, P.O.B. 1016, Yu-41001 Zagreb, Croatia, Yugoslavia

(Received May 6, revised September 10/Accepted September 11, 1987)

The concept of *composition* as the counterpart to *partition* is introduced and advocated for the discussion of molecular properties. In the partition approach an observable (experimental) quantity is fragmented into contributions which are non-observable but which hopefully maintain constancy for fragments (bonds) in similar environments and thus facilitate comparison of data. With the composition as an approach the role of “fragments” and “whole” are reversed: one starts with a collection of observable fragment properties (e.g., atomic chemical shifts of NMR spectra) and then constructs an abstract non-observable quantity representing the collection of fragments as a “whole”. If a so-derived quantity for different molecules shows some regularity, the initial loss of information in condensation of independent fragment data is compensated by insight into novel structural correlations. The approach is illustrated first by ordering isomers (e.g., nonanes C₉H₂₀) with respect to their content of special graph invariants p_2 and p_3 (numbers of paths of length two and length three, respectively) and then showing that the constructed global quantity derived from individual carbon-13 NMR chemical shifts shows a regular variation with p_2 and p_3 , very similar to isomeric variations of numerous thermodynamic properties of nonanes. Subsequently it is outlined how the difference ($p_2 - p_3$) leads to a correlation for mean carbon-13 chemical shifts in octanes and nonanes, taken as an illustration for the approach. It is expected that the outlined approach opens new avenues for data reduction and the search for structure-property correlations.

Key words: Chemical graph theory — Composition — Alkanes — Molecular chemical shift

* Dedicated to Professor J. Koutecký on the occasion of his 65th birthday

** Operated by Iowa State University for the US Department of Energy under Contract No. W-7405-ENG-82. This work was supported in part by the Office of the Director

1. Introduction

The search for regularities in data, empirical correlations, bond additivities and alternative data reduction schemes have been the subject of continuous interest in chemistry for a long time [1, 2]. The central role which such efforts play in science has been generally recognized. "For every science," wrote Max Planck [3], "not even excluding mathematics, is to some extent the result of observation, whether the subject be natural or intellectual. The chief problem in every science is that of endeavouring to arrange and collate the numerous individual observations and details which present themselves, in order that they may become part of one comprehensive picture." There are numerous avenues which could be followed in an effort to "arrange and collate the numerous individual observations" just because there are numerous different types of observations to start with. Mendeleev's Periodic System of Elements is one such useful mode of "arranging" data, when one considers atomic properties, whereas the Schrödinger equation as a model offers a frame for sorting out observations relating to atomic or molecular spectra. In both cases the basis for regularities is in the electronic structure and symmetry that those atomic and molecular system possess. But there are numerous molecular properties and numerous regularities that cannot be rationalized from the first principles governing the motions of electrons and atomic nuclei at the molecular level. The model, such as the Schrödinger equation, becomes a source of new data such as total molecular binding energy, molecular ionization potential, molecular geometry, dipole moment, etc. Comparisons between molecules have then led to new concepts, such as resonance energy as a measure of additional stabilization of polycyclic fused conjugated hydrocarbons in comparison to an acyclic (olefinic) reference structure, or bond orders as a measure of "compactness" of a bond and its length. The list of useful and used concepts is impressive: potential curves, hybrids and hybridization, molecular orbitals including HOMO and LUMO orbitals, frontier orbitals and general nodal properties of orbitals, equivalent orbitals and localization procedures, etc. All these, as is known, are non-observable quantities and the question is if such use is legitimate and how it is to be justified.

One uses non-observable quantities as intermediate parameters, parameters that serve to facilitate a comparison between observables. Their role is similar to that of dummy variables in evaluation of integrals. The selection of coordinates, for example, does not change the value of an integral which may represent an observable quantity (e.g. a volume), but a judicious selection of coordinates may make calculation possible, even simple, which otherwise may fail to be resolved. Hence, non-observable quantities used in structural chemistry are not only legitimate but evidently are a necessity. Strictly they serve in the purpose of *comparison* of data pertaining to different molecules or various components of the same molecule.

2. Partition techniques

One of the models used in comparison of molecular properties, particularly in the study of isomeric variations is *fragment additivity* [4]. Atomic additivity and

bond additivity are the simplest illustrations, perhaps the next simplest model involves circuit additivity, etc. By assuming an additivity one is partitioning the observable property into components (bond terms for instance) which are not observables. Intuitively one justifies this process by virtue of constancy (or approximate constancy) of representative terms which allows transferability of the parameters among structurally related molecules.

As an illustration of partitioning of a molecular property into atomic, bond and pair of bonds contribution consider diamagnetic susceptibilities of saturated organic molecules. Hameka [5] developed a general MO model for this problem assuming the following: 1) All bonds are localized; 2) all bonds are identical (CC and CH bonds, respectively); and 3) all bond lengths are constant. These the basic underlying assumptions imply that details of molecular geometry are not relevant for the model, and as recognized subsequently, the MO formalism of Hameka can be cast fully in an equivalent elegant graph theoretical form [6]. By adopting basic notions from graph theory the analysis is reduced to the enumeration of selected graph invariants. A different partition arises in the analysis of resonance energies of conjugated hydrocarbons. Here independently calculated resonance energies (parametrized against a particular SCF π -MO scheme [7]) is found to be circuit additive. Not all circuits are of interest; only those circuits in individual Kekulé valence structures which can be associated with an alternation of CC single and double bonds play a role [8-10]. The result represents an interesting novel connection between concepts of MO and VB theory, and the underlying concept of conjugated circuits has been found important for the clarification of other properties of conjugated hydrocarbons [11-21].

3. Composition as a method for data reduction

Partition as a method in searching for additives is well and widely known. It is then surprising to realize that the conceptually related method, composition, in which one reverses the role of "whole" and "fragments" and used observable quantities as components from which an abstract (non-observable) quantity is constructed has not been recognized as a potentially important technique for data reduction. Because the process can be visualized as that inverse to partition we need not dwell on the legality of the method, but the same caution that pertains to partition technique applies also here. In particular a justification for an application will mostly depend on what use such an approach offers in discussion of molecular properties. The distinction between the two methods, partitioning and composition may in some situations be a matter of emphasis, e.g., in situations when both "whole" and "fragments" are equally accessible from experiment or calculation. A trivial illustration is the evaluation of molecular weight as a sum of atomic weights [22]. More interesting cases are when the fragment property is easily recognized, while the corresponding molecular concept is non-existent. An example is the carbon-13 chemical shifts in NMR spectra of alkanes. Here we have only the atomic property; there is simply no "molecular" chemical shift at least until now! It is helpful to make a comparison with a typical partitioning to recognize the reverse situation there. For example boiling point

is a molecular property, there is no "fragment" boiling point (here rather than atomic we would have, at least as dominant a bond contribution). But by partitioning the property (i.e. assuming there is some bond or/and other additivity fragments) we associate with individual terms corresponding boiling point fraction. Hence one could in principle in the case of alkanes speak of "bond boiling point" meaning the contribution to the additivity for the molecular boiling points arising from the bond in question. In the case of boiling points such vocabulary was never developed, appearing somewhat awkward, but we do partition in bond contributions many other molecular properties and speak of bond energies, bond dipoles, etc. despite the fact that these are as hypothetical as bond boiling points would be (except in diatomic molecules). It follows then, in reversing the role of "partition" and "composition" that indeed we could speak of "molecular chemical shift" as a chemical shift hypothetical value associated with a molecule as a whole and based on atomic contributions. Or we could find a more palatable name for such abstract construction. But the final judgement will depend on what use can come from such a construction. Can such molecular properties in some way represent data reduction, can they tell us something more about the structures involved that we did not deduce from collection of independent atomic data? The answer is "yes", at least for the case which we have selected as an illustration.

4. Mean atomic carbon-13 NMR chemical shifts in alkanes

The suggestion that carbon-13 chemical shifts in alkanes may provide an illustration of the construction of a "molecular" property has already been made and illustrated on octane isomers [24]. It has not been emphasized there that this is a case of the general *composition* technique; merely sum of carbon-13 chemical shifts for isomers of octane were constructed and then it was argued that the resulting numbers show a behavior typical of other thermodynamic molecular properties, i.e. a regular variation among the isomers. As an observation the finding is novel: 18 independent sets of data have suddenly been found to show a particular regularity; a regular decrease or increase of the chemical shift sums when structures are arranged on a grid specially constructed for comparison of molecular properties. As one moves across the grid vertically or horizontally one notices a kind of periodic variation in the property reminiscent of variations in rows and columns of the Periodic Table of Elements. In fact the grid can be viewed alternatively as a table, which can legitimately be referred to as a Periodic Table of Isomers [25–28] (see Figs. 1, 2).

We will here illustrate the same regularities for nonane isomers in order to demonstrate a more general nature of the property. The 18 isomers of octane already provide convincing evidence for the regularity, but since here we consider *trends*, not correlations based on numerical comparisons (of predicted and actual) values, one has to be even more cautious not to give significance to possible chance factors. With the consideration of similar quantities for nonane isomers we have more than doubled the number of observations and practically eliminated an objection that the regularity observed may be fortuitous.

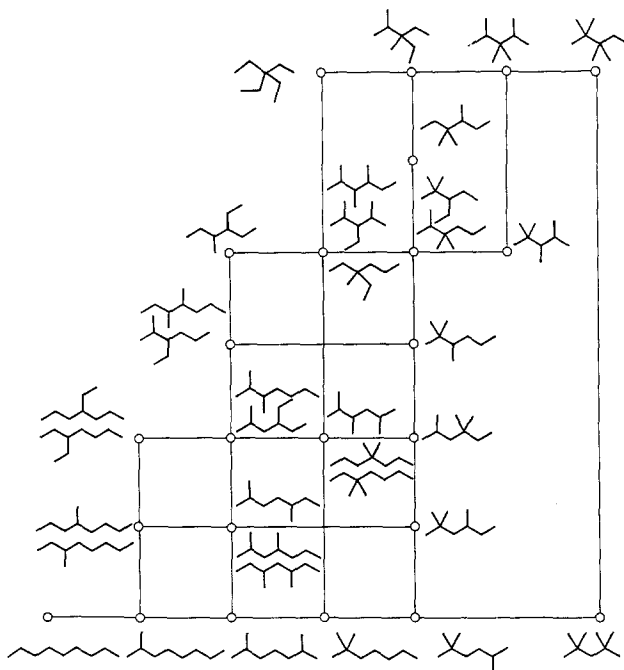


Fig. 1. (p_2, p_3) grid for nonane isomers C_9H_{20}

n						
2-M	3-M 4-M	3-E 4-E				
2,6-MM	2,4-MM 2,5-MM 3,5-MM	2,3-MM 2,4-ME	2,3-ME 3,4-MM	3,4-ME		
2,2-MM	∅	2,3,5-MMM 4,4-MM 3,3-MM	∅	2,3,4-MEM 2,3,4-MMM 3,3-ME		3,3-EE
2,2,5-MMM	2,2,4-MMM	2,4,4-MMM	2,2,3-MMM	2,3,3-MMM 2,2,3-MME	3,3,4-MMM	2,3,3-MME
				2,2,3,4-MMMM		2,3,3,4-MMMM
						2,2,3,3-MMMM
2,2,4,4-MMMM						

Fig. 2. An alternative representation of the ordering of isomers based on count of graph invariants p_2 and p_3

Before outlining the *ordering* rule for the structures a comment regarding the interpretation of the constructed composed quantities is in place. Instead of considering a collection of carbon-13 NMR chemical shifts for a molecule we propose to simply view these as "atomic" contributions and proceed to construct the corresponding sum. Clearly any atomic property can in this way be reduced to a single item and the process is accompanied with a *loss* of information. Unless we can subsequently obtain some new insight the proposition would be at best inconsiderate. In Table 1 we have listed the carbon-13 sums (i.e. we simply added chemical shifts of all carbon atoms in each molecule). Already this table shows some intriguing details. Observe the relatively large spread of the values for carbon-13 sums, from the value of 173.9 for 3,3-diethylpentane to 321.1 for 2,2,4,4-tetramethylpentane:



Table 1. Nonane isomers for which carbon-13 NMR chemical shifts were reported ordered relative to the magnitude of the corresponding chemical shift sum or mean carbon-13 chemical shift. Observe the parallelism with the difference $p_2 - p_3$, p_2 and p_3 representing paths of length two and length three, respectively. Even with the exception of 2,4-dimethyl-3-ethyl-hexane which appears in two diastereomers one of which (but not the average shown) would place the isomer above 2,3-dimethyl-heptane, in agreement with other structures

Nonane isomer	(p_2, p_3)	$p_2 - p_3$	Mean	Exp.-calc. difference
3,3-diethyl	10, 12	-2	19.32	-1.48
3-ethyl	8, 8	0	23.53	-0.17
2,3,3,4-tetramethyl	12, 12	0	23.82	0.12
3,4-dimethyl	9, 9	0	24.03	0.33
2,3-dimethyl	9, 8	1	25.01	-0.13
<i>n</i> -	7, 6	1	25.04	-0.01
3-methyl	8, 7	1	25.21	0.07
2,3,3-trimethyl	11, 10	1	25.24	-0.10
2,4-dimethyl-3-ethyl	10, 10	0	25.38	1.68
4-methyl	8, 7	1	25.40	0.26
3,5-dimethyl	9, 8	1	25.42	0.28
2,2,3,3-tetramethyl	13, 12	1	25.46	0.32
3,3-dimethyl	10, 8	2	25.84	-0.75
2,3,5-trimethyl	10, 8	2	26.30	-0.29
2-methyl	8, 6	2	26.36	-0.23
2,5-dimethyl	9, 7	2	26.49	-0.10
4,4-dimethyl	10, 8	2	26.76	0.17
2,4-dimethyl	9, 7	2	26.77	0.18
2,2,3,4-tetramethyl	12, 10	2	27.48	0.89
2,6-dimethyl	9, 6	3	27.78	-0.26
2,2-dimethyl	10, 6	4	28.47	-1.01
2,2,4-trimethyl	11, 7	4	29.74	0.26
2,2,5-trimethyl	11, 6	5	29.76	-1.17
2,2,4,4-tetramethyl	13, 6	7	34.68	0.86

One wonders why there are such great differences and what is the meaning of the difference? What is particular about the above two structures that they show such widely different sums? Chemical shift sums are of no interest as isolated molecular parameters, they become significant precisely at the time one makes *comparisons* among molecules. From comparisons we expect to *gain* in the information and compensate for the loss that occurred during the reduction process of the new parameter. Chemical shift or chemical shielding constants measure the decrease of an external magnetic field at the location of each particular nucleus. The induced fields are proportional to the external field (with negative proportionality constants). Chemical shift sums then would correspond to a view of a molecule as nondescript point surrounded by electron clouds, obviously of different *shape*, because the corresponding carbon-13 chemical shift sums differ so much from case to case.

The numerical values are based on work of Grant and Paul [29] and the subsequent compilation of Lindeman and Adams [30].

If the derived chemical shift sums are normalized by dividing the values with the number of carbon atoms in the molecule one apparently derives an average or mean carbon-13 chemical shift for individual alkanes. Because carbon-13 chemical shifts are an atomic property, it seems more agreeable to refer to the composition when suitably normalized as a “mean atomic” property. An additional advantage of the use of “mean carbon-13 chemical shift” over “chemical-shift sum” is that the former facilitates comparisons between molecules of different size, for instance between octane and nonane isomers. One finds for instance on comparison of the values for *n*-octane and *n*-nonane that with the increase in the size of the molecule also a small but significant increase in the mean carbon-13 chemical shift occurs (24.45 ppm and 25.04 ppm, respectively are the values for normal C_8H_{18} and C_9H_{20}). However, in contrast to the previously mentioned boiling points and their fragmentation into bond contributions, viewing the chemical shift sums as a “molecular” property is not without some merit.

5. Ordering of structures using graph invariants

We selected the number p_i of paths of length i (the metric on graphs is defined by the count of intervening bonds) as the basic graph invariants *because past experience* [25–28] indicated that these particular quantities play a useful role in comparison of molecular properties. A path is defined as a sequence of consecutive bonds; p_1 simply represents the bonds, p_2 enumerates chains of two adjacent bonds and so on. In alkane isomers differentiation among structures starts with p_2 and continues with use of longer paths. The results can conveniently be presented as a sequence: $p_0, p_1, p_2, p_3, p_4, \dots$. Path enumerations for nonane’s 35 isomers have been reported. If one truncates such path sequences to only the pair p_2, p_3 the corresponding numbers can be interpreted as coordinates on a rectangular grid. It follows that most isomers have different coordinates, as illustrated in Fig. 1. Because the more distant neighbors in molecular additivity

schemes play a decreasing role, higher terms which make a better discrimination among structures have been excluded. Several pairs of isomers (and even triplets) have the same characterization, but while this would be a disadvantage if codes are used for documentation-oriented purposes, here the occurrence of coincidences are not necessarily troublesome. In fact they may be even an advantage, because *different* molecules may have and do have occasionally the *same* or very close values for selected properties. The actual enumeration of paths for acyclic structures is rather straightforward [31] as illustrated in Fig. 3 for paths of length 1–4 in 2,3,3,4-tetramethylpentane initiated from atom labeled **a**.

For the four non-equivalent carbon atoms of this molecule one obtains the following sequences p_1, p_2, p_3, p_4 :

a: 1, 2, 3, 2

b: 3, 3, 2, 0

c: 4, 4, 0, 0

d: 1, 3, 4, 0

which when added with appropriate symmetry weights due to the presence of equivalent classes of atoms (there being four atoms **a**, two atoms **b** and **d**, and one atom **c** in this case) will give twice the number of all corresponding paths of respective length for the molecule as a whole, twice, because each path is counted twice, once for each end atom. Hence one easily deduces the path sequence for 2,3,3,4-tetramethylpentane as: 8, 12, 12, 4 and the (p_2, p_3) coordinates for the isomer being (12, 12).

Once the structures (isomers) have been ordered (Fig. 1) one *replaces* the individual structures with a selected structural property. If the resulting magni-

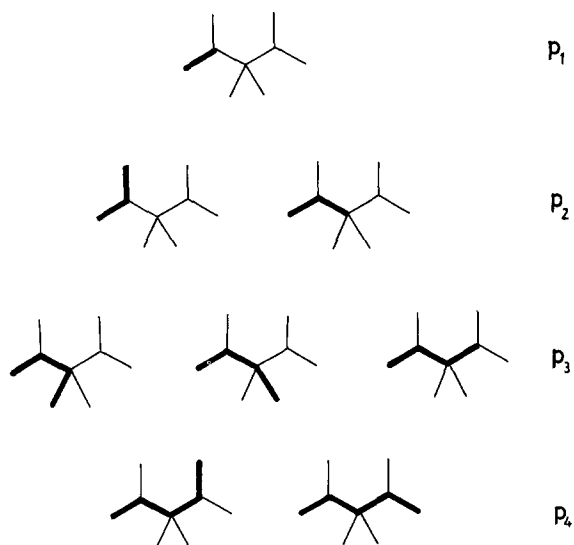


Fig. 3. Illustration of enumeration of paths p_i of different length for an isomer of nonane

tudes show a pattern one can rightly deduce that the rules governing the ordering of structures also govern the ordering of the particular property. In Fig. 4 we show the resulting relative magnitudes for the mean carbon-13 chemical shift in nonane isomers.

A regular increase in the mean chemical shift (a molecular property!) is to be observed as we go from left to right and as we go from top to bottom.

Figure 4 enables *novel* insights concerning NMR data for a collection of molecules. The derived pattern is not random, the ordering reflects inherent structural differences among isomers. It would be somewhat imprecise to associate the trends with molecular "shape", for one thing because the concept of "shape" has yet to be better defined. From Fig. 1 we see that moving from left to right can be intuitively associated with decrease in elongation and increase in "sphericity" of a molecule. But so is, to a lesser degree the case with vertical comparisons of structures, where the trend is reversed: more "elongated" structures (at the bottom of the table) have larger not smaller mean carbon-13 chemical shifts. This points to a need for *two* "shape" parameters such as are the parameters p_2 and p_3 even if their interpretation as "shape" parameters is not fully attainable.

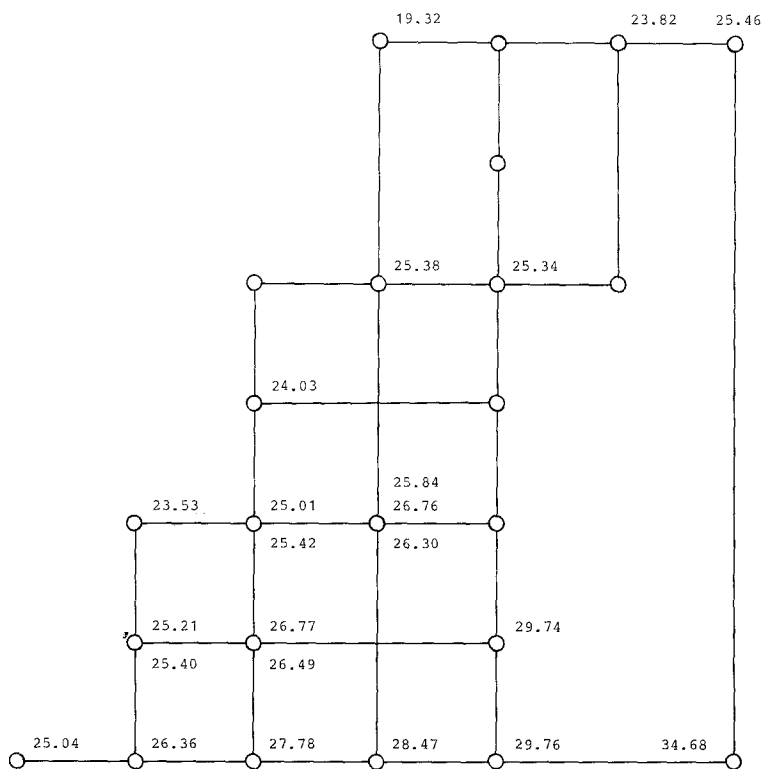


Fig. 4. Trends in the relative magnitudes of mean carbon-13 chemical shifts for nonane isomers depicted by replacing the structure in the grid on Fig. 1 by the numerical value for the property

6. A correlation between mean chemical shifts and a structural parameter

While trends shown in Fig. 4 justify the concept of average or mean chemical shift as a valuable molecular characteristic, objections may be raised that by its very nature such an approach is necessarily qualitative, hence inferior to quantitative structure-property correlations and quantitative additivities. One of the uses of the ordering of structures as shown in Fig. 1 is to lead us towards quantitative correlations. Figure 4 suggests that there may be a correlation between the structure and the composed property by the fact that trends show regularities. We know that similar regularities in the relative trends have been observed for numerous thermodynamic properties for which also correlations and additivities were reported. So one anticipates a correlation between the structure (yet to be adequately characterized) and this constructed molecular parameter. But instead of trial and error one can make some educated guesses as to what may be the important factors to be considered. A closer look at Fig. 4 reveals that structures along a diagonal parallel to the one connecting the extreme "shapes" of *n*-nonane (a linear chain) and 2,2,3,3-tetramethylpentane (a more "spherical" structure) have a near constant mean carbon-13 chemical shift, the values for the two mentioned isomers being 25.04 ppm and 25.46 ppm respectively. This alone indicates that the intuitive notion of a "shape" here is not the important structural factor and serves as a warning for uncritical extension of familiar concepts to new domains. The observation however allows an educated guess that the parameter ($p_2 - p_3$), which figuratively represents the diagonal direction of the grid of Fig. 4, may provide a basis for useful correlation. In Fig. 5 we have plotted the mean carbon-13 shifts for octane (marked as ●) and nonane (marked as ○) isomers.

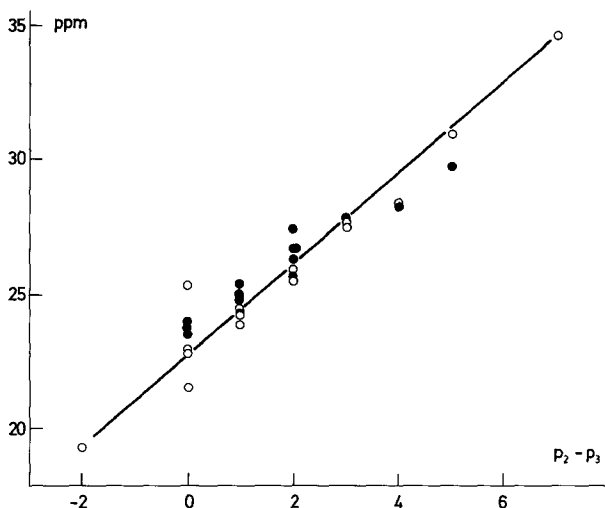


Fig. 5. A correlation between the mean carbon-13 chemical shifts of octane C_8H_{18} and nonane C_9H_{20} isomers (marked as ● and ○, respectively) and the difference in the path numbers of length two and three ($p_2 - p_3$)

Figure 5 represents a histogram, since one coordinate is an integer, nevertheless the correlation produced is quite satisfactory. To appreciate the significance one should recollect that $(p_2 - p_3)$ is a fixed structural parameter completely determined by molecular connectivity yet all data on octanes (18 points) and all available data on nonanes (24 points out of 35 possible) have been included in the diagram. Recently, a few additional carbon-13 chemical shifts were reported [35] but a comparison of the results for 3,4-dimethylhexane appears to give slightly higher values for some chemical shifts than as reported in the work of Lindeman and Adams [30] (mean chemical shift being 23.96 ppm as compared with 23.08 ppm), and regardless which set of values is more reliable their inclusion would possibly require revision of other points or scaling down of the results for additional ones (one octane and two nonane isomers) that it was thought unwarranted. Nevertheless, the work of Lachance, Brownstein and Eastham [35] which was mostly devoted to the problem of the identification of aliphatic hydrocarbons containing multiple asymmetric centres, indicates the associated variations in chemical shifts and chemical-shift sums among diastereomers to be of the order of the dispersion of the points in Fig. 5. For example in case of 2,3,4-trimethylhexane the mean carbon-13 chemical shifts for the two diastereomers are 24.0 and 25.40, a spread of 1.40 ppm which typifies the spread of points in Fig. 5 around the correlation line.

Some dispersion of points in any correlation is to be expected, even more so when one of the "coordinates" is by its nature restricted to integers. Perhaps what is remarkable about Fig. 5 is the observation of practically identical mean carbon shift values for apparently very different structural forms. For example, the pairs:

- (a) *n*-nonane and 2,3-dimethylheptane;
- (b) 4-methyloctane and 3,5-dimethylheptane;
- (c) 2,2,5-trimethylhexane and 2,2,4-trimethylhexane, etc.

These coincidences are expected from same values of $(p_2 - p_3)$ but within ± 1.5 ppm. The agreements within ± 0.05 may be fortuitous, but may also signify coincidences in additional graph invariants governing minor variations within a class of isomers with the same value of $(p_2 - p_3)$. This remains to be seen provided sufficient number of cases become available to merit such considerations. Finally we could add one advantage of the outlined analysis: it is not sensitive to details of the assignment of NMR signal lines, in fact one need not know the *assignment* of the lines at all in order to complete the above analysis!

7. Concluding remarks

The composition described provides some insight into isomeric variations of the relative magnitudes of average chemical shifts in alkanes and shows that the regularities found are similar to those also found for the same compounds when their thermodynamic properties were studied. The practitioner may nevertheless be disenchanted, expecting some benefits for future work. Well, the mean carbon-13 chemical shifts in alkanes (and similar relations may exist for families of

structurally related compounds) may help in suggesting the location of missing peaks in an isomer if similar data are available for compounds in the neighboring position on the grid or Isomer Table. The greater value of the approach is on the conceptual level. Since the pioneering work of Grant and Paul [29] on the significance and role of molecular connectivity, i.e., number of neighbors for acyclic structures, interest remained in exploring the role of these factors in more detail. Here we see emerging, after more than 20 years of accumulation of data, the possible role of yet another structural factor, the difference in the number of paths of length two and length three: $(p_2 - p_3)$. It need not be immediately clear why $(p_2 - p_3)$ is relevant here, but the first thing to observe is that it is the critical variable. Application of graph theory to chemistry, and then to structure-activity correlations has led to the emergence of several critical graph-theoretical indices, such as the count of non-adjacent pairs [36], the connectivity index [37], centric index [38], and even a purely topological parameter: the cardinality of associated graph topology [39], and many others [34, 40]. These have been found very useful and suitable for discussion of isomeric variations, particularly (but not exclusively) in alkanes. All these indices have an intriguing feature: they work surprisingly well and yet leave some mystique as to why this is so. The same has been true also with the graph-theoretical indices introduced by Wiener [41, 42] in his pioneering work on a structural approach to correlations, the so-called Wiener numbers W and P . In his own words: “ W is the sum of all the distances, in terms of carbon-carbon bonds, between the members of all pairs of carbon atoms in the molecule; P is the number of pairs of carbon atoms separated by three bonds” [49]. In our symbolism P is in fact p_3 while W is the sum of all p_i values – remarkable constructions, immediately recognized by Platt [42, 43] as fundamental for the study of structure-property relations, but largely ignored by other researchers until the recent revival of chemical graph theory. We may not yet have all answers as to why some paths are of interest here, and some others there, but by accumulating data and correlations a simple picture is likely to emerge. Numerous comparative studies of various topological indices [40, 45–49] have already been useful in clarifying some of the differences among them as well as offering insights in some of their “mystique”. Additional application may only facilitate the continuing process for alternative data reductions, model constructions and understanding of quantitative structure-property relationships. Here “understanding” is to be interpreted as the expression of less familiar concepts in terms of more transparent ones, and frequently easily computed ones. Path numbers p_i may well represent one of the basic concepts in such efforts.

Finally one may consider as an alternative a statistical analysis. That $p_2 - p_3$ is useful parameter could conceivably be deduced after one performs a statistical factor analysis using all p_i as basis, thereby obtaining statistical confidence measures. In addition, one could create arrays of random numbers and check whether it is possible to obtain reasonable weights from a factor analysis on this basis. Whilst such suggestions may be of interest they involve some limitations that are not apparent. First, the idea of considering linear combinations of path numbers came from the graph theoretical analysis. Prior to the ordering of isomers

we had no notion that path numbers may provide the basis for the correlation, the least that combinations like $p_2 - p_3$ will play a dominant role. However, even if the statistical analysis would point to $p_2 - p_3$ as the dominant structural parameter we would be in the dark as to why this is the case. The case of NMR shifts in alkanes well illustrates the difference between the statistical approaches such as factor analysis and graph theoretical approaches where one can identify the critical structural parameters from observations rather than from computer read-outs of statistical packages, that at best give answers, but cannot anticipate them.

Acknowledgements. We thank the referees for useful comments.

References and notes

1. Auwers KV (1915) *Leibig Ann.* 410:287
2. Fajans K: *Ber* (1920) 52:643
3. Planck M (1960) *Survey of physical theory.* Dover, New York
4. Klein DJ (1986) *Int J Quantum Chem: Quantum Chem Symp* 20:153
5. Hameka HF (1961) *J Chem Phys* 34:1966
6. Randić M (1978) *Chem Phys Lett* 53:602
7. Dewar MJS, de Llano C (1969) *J Am Chem Soc* 91:789
8. Randić M (1976) *Chem Phys Lett* 38:68
9. Trinajstić N (1983) *Chemical graph theory, vol II, chap 3.* CRC Press, Boca Raton
10. Randić M, Trinajstić N (1984) *J Am Chem Soc* 106:4428
11. Randić M (1977) *J Am Chem Soc* 99:444
12. Randić M (1977) *Tetrahedron* 33:1905
13. Randić M (1977) *Mol Phys* 34:849
14. Gutman I, Randić M (1979) *Chem Phys* 41:265
15. Randić M (1980) *Int J Quantum Chem* 17:549
16. Randić M (1980) *Pure Appl Chem* 52:1587
17. Klein DJ, Schmalz TG, Hite GE, Seitz WA (1986) *J Am Chem Soc* 108:1301
18. Schmalz TG, Seitz WA, Klein DJ, Hite GE (1986) *Chem Phys Lett* 130:203
19. Klein DJ, Seitz WA, Schmalz TG (1986) *Nature* 323:703
20. Randić M, Nikolić S, Trinajstić N (1987) *Gazz Chim Ital* 117:69
21. Randić M, Nikolić S, Trinajstić N: In: King, RE, Ronvray DH (eds) *Applications of topology and graph theory to chemistry.* Elsevier, Amsterdam, in press
22. Even the partition of a given quantity into atomic contributions may be far from trivial in some instances. An example is provided by Blinder [23] in a paper on additivity rules in molecular thermodynamics in his discussion of translational entropy and the question of how would an atom "know" how many other atoms a molecule has. As shown there concepts of graph theory help to resolve the dilemma
23. Blinder SM (1975) *J Am Chem Soc* 97:978
24. Randić M (1980) *J Magn Res* 39:413
25. Randić M, Wilkins CL (1979) *J Phys Chem* 83:1525
26. Randić M, Wilkins CL (1979) *Chem Phys Lett* 63:332
27. Randić M, Wilkins CL (1980) *Int J Quantum Chem* 18:1005
28. Randić M, Trinajstić N (1982) *Math Chem (Mülheim/Ruhr)* 13:241
29. Grant DM, Paul EG (1964) *J Am Chem Soc* 86:2984
30. Lindeman LP, Adams JQ (1971) *Anal Chem* 4:1245
32. Nevertheless if a molecule has a dozen atoms the enumeration becomes not only tedious but error-prone. In polycyclic structures enumeration of paths becomes prohibitive already for structures having more than one cycle even if the number of atoms is relatively small. One has

- to resort to computer assistance. Program has been outlined (implemented in PL/1, BASIC and FORTRAN) [32, 33]. However, even when the computer enumeration of paths is carried out for polycyclic systems, there are some unavoidable difficulties (34)
32. Randić M, Brissey GM, Spencer RB, Wilkins CL (1979) *Comput Chem* 3:5
 33. Randić M, Brissey GM, Spencer RB, Wilkins CL (1980) *Comput Chem* 4:27
 34. Knop JV, Müller WR, Szymanski K, Trinajstić N (1986) *Int J Quantum Chem: Quantum Chem Symp* 20:173
 35. Lachance P, Brownstein S, Eastman AM (1979) *Can J Chem* 57:367
 36. Hosoya H (1971) *Bull Chem Soc Japan* 44:2332
 37. Randić M (1975) *J Am Chem Soc* 97:6609
 38. Balaban AT (1979) *Theor Chim Acta* 53:355
 39. Merrifield RE, Simmons HE (1980) *Theor Chim Acta* 55:55
 40. Bonchev D, Trinajstić N (1977) *J Chem Phys* 67:4517
 41. Wiener H (1947) *J Am Chem Soc* 69:17
 42. Wiener H (1947) *J Am Chem Soc* 69:2363
 43. Platt JR (1947) *J Chem Phys* 15:419
 44. Platt JR (1952) *J Phys Chem* 56:328
 45. Balaban AT, Motoc I (1979) *Math Chem (Mülheim/Ruhr)* 5:197
 46. Balaban AT, Chiriac I, Simon Z (1980) *Lect Notes Chem*, vol 15. Springer, Berlin Heidelberg New York
 47. Bonchev D, Mekenyan O, Trinajstić N (1981) *J Comput Chem* 2:127
 48. Razinger M, Chretien JR, Dubois J-E (1985) *J Chem Inf Comput Sci* 25:23
 49. Trinajstić N, Randić M, Klein DJ (1986) *Acta Pharm Jugosl* 36:267