CROATICA CHEMICA ACTA CCACAA **66** (3–4) 411–434 (1993)

ISSN 0011-1643
CCA–2150

Feature article

Graph-Theoretical Correlations – – Artifacts or Facts?

*Milan Randić**

Department of Mathematics and Computer Science,
Drake University, Des Moines, Iowa 50311–4505, USA

Zlatko Mihalić

Faculty of Science, University of Zagreb, HR–41000 Zagreb, Croatia

Sonja Nikolić and Nenad Trinajstić

The Rugjer Bošković Institute, HR–41001 Zagreb, Croatia

Received September 21, 1993

> »...the Fact is the basis, the foundation;
> Imagination the building material; the
> Hypothesis the ground plan to be tested;
> the Truth or Reality is the building...«
> J. H. van't Hoff[1]

Structure-property correlations and use of various auxiliary modes of representing them are critically examined. Several rules are suggested for the evaluation of alternative schemes and their potential application. Distinction is made between empirical, theoretical and structural correlations, which helps settle the question of whether graph-theoretical correlations are artifacts or facts.

---

* Author to whom correspondence should be addressed.

INTRODUCTION

The search for empirical correlations, bond additivities and regularities in isomeric variation has been a subject of continuous interest in chemistry for a long time.[2] Such efforts represent an attempt to reduce data, which is, and remains, one of the central problems in science: »*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 endeavoring to arrange and collate the numerous individual observations and details which present themselves, in order that they may become part of one comprehensive picture.*« Interest in correlations, additivities and isomeric variations originates, one may say, in the inherent challenge of developing an adequate model for the chemical structure that would reproduce available data to a satisfactory accuracy. There are, however, no rules to follow how to construct a model, what components are preferable and how the models are to be evaluated. In the case of the Schrödinger equation, the variational principle allows one to compare and select a better solution, but in the case of molecular additivities no similar procedure exists. As discussed by Smolenskii[4] and elaborated by Gordon and Kennedy,[5] if a sufficiently large pool of components is selected, any physical property can be reproduced accurately. On the other hand, even when one refrains from proliferating empirical or structural parameters and arrives at a model with a limited number of parameters (say a dozen) for a particular molecular property which produces acceptable predictions, this does not necessarily mean that the model is optimal. Perhaps, another model with half as many parameters and somewhat reduced accuracy might much better represent the molecular structure because it also applies to *other* qualified properties. To paraphrase Albert Einstein: »*The model should be as simple as possible, but not simpler!*« The problem is in identifying such models – the topic that we address here. The accuracy of the prediction cannot be the only and ultimate criterion for judging models. Elaborate models can frequently produce accurate predictions, but this does not mean that a simpler model cannot do the same or that, since they predict the considered property accurately, they necessarily represent real objects faithfully. For example, the geocentric planetary model was capable of accurate predictions of eclipses (based on the complicated systems of epicycles) but it was devoid of reality.

PROPERTY–PROPERTY CORRELATIONS

Construction of a property-property correlation (and extension of such considerations to a collection of properties) is the most elementary data reduction because it does not presume any model for the structure. In Figure 1, we illustrate such a correlation: plot of the liquid densities against the refraction indices for octanes.

We intuitively consider both properties as volume-dependent, *i.e.*, capable of being reduced to the same structural element(s). However, strictly speaking, all that we can deduce from Figure 1 is that *both* properties depend on the *same* structural factors, if indeed they can be shown to be »structural« properties. A structural property by definition is a property that can be expressed in terms of well-defined structural concepts, such as atoms, valencies, bonds, bond types, number of the nearest neighbors, number of the next-nearest neighbors, rings, vertices, edges, paths, walks, *etc.*

In Figure 2, we illustrate another property-property correlation: a plot of the boiling points against enthalpies of atomization for octanes.

Again a valid correlation is demonstrated, but the somewhat larger scatter of points suggests that, besides the common dominant structural factor(s), the two properties

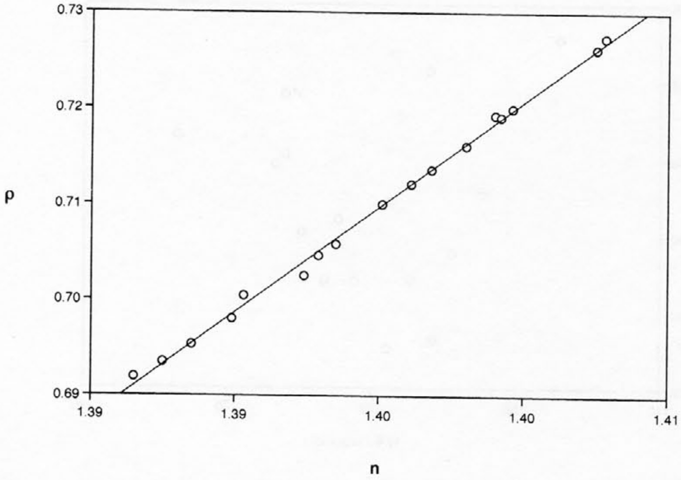

Fig. 1. A plot of the liquid densities (*p*) *vs.* the refraction indices (n) of octanes.

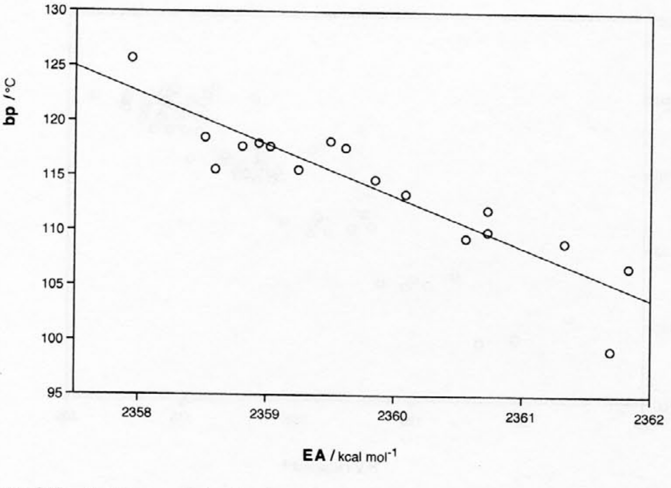

Fig. 2. A plot of the boiling points (bp in °C) *vs.* the enthalpies of atomization (*EA* in kcal/mol) of octanes.

may also depend on (an)other factor(s). Finally, in Figure 3 we show a plot of two properties without any apparent correlation, that is, molar volumes against heats of vaporization for octanes. However, if we plot molar volumes against heats of vaporization for alkanes from pentanes to nonanes (see Figure 4), then some limited correlation between these two properties emerges.

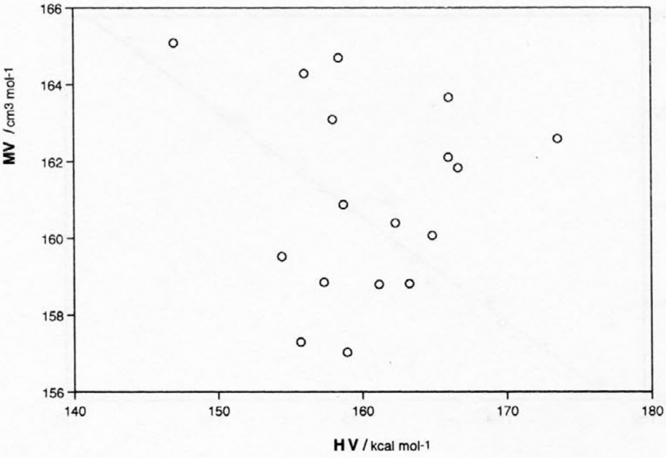

Fig. 3. A plot of the molar volumes (*MV* in cm^3/mol at 20 °C) *vs.* the heats of vaporization (*HV* in kcal/mol at 25 °C) for octanes.

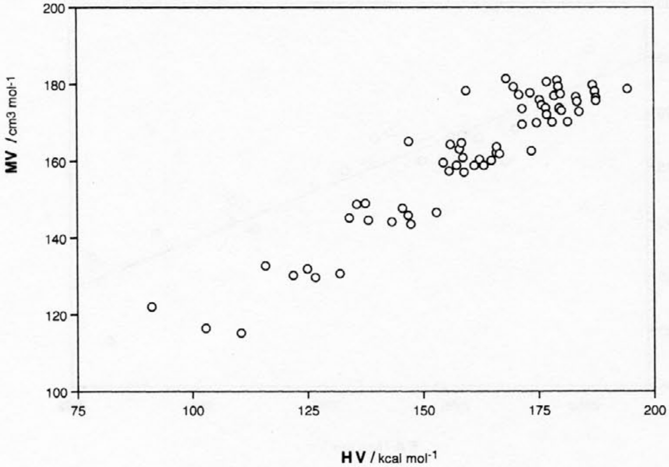

Fig. 4. A plot of the molar volumes (*MV* in cm^3/mol at 20 °C) *vs.* the heats of vaporization (*HV* in kcal/mol at 25 °C) for alkanes from pentanes to nonanes.

It is clear from Figure 4 that two properties considered depend on two or more structural parameters. Usually, the minimum number of parameters that would suffice can be determined using discriminant analysis. In fact, a generalization of property-property correlation is a multivariate technique[6,7] applied to a collection of properties. For example, some years ago Cramer[8] considered the following properties for a diverse collection of compounds: aqueous solvation, octanol-water partition, molar refractivity, boiling points, molar volumes and heats of vaporization. Compounds studied are found to correlate their properties among themselves as a collection, but not so well on a pairwise basis. A factor analysis approach,[9] to speak figuratively, represents a »blind« man's approach to the correlation problem, introducing no bias whatsoever. This is a good side of the method, which, when combined in a study of clustering of similar compounds (which can be accomplished, as it has been recognized,[10] at various levels of classification), offers a powerful tool for molecular studies. But all that can be correctly deduced from the work of Cramer is that the six molecular properties mentioned depend on the *same* structural factors. What these factors may be remains unknown and speculations on the possible role of various structural factors have no basis in the obtained results. The situation is analogous to the case of a simple property-property correlation, such as the one shown in Figure 1 which, strictly speaking, only shows that the two properties considered have a *common* structural basis. It is outside the domain of *empirical* schemes (a correlation or a factor analysis) to reach conclusion on *structural* relations when such schemes do not use structural concepts. If some structural concepts have been employed or implied, the approach no longer represents a purely empirical method.

The step from a correlation to an additivity (or some generalized expressions) requires development of an adequate partitioning of the properties which, in turn, is assisted by an appropriate model for a structure. Hence, the study of empirical correlations may result in evolution of valuable structural models. We will later review a few of such successful schemes, but some caution is in place even at this first step in data reduction. The fact that two properties lead to a good correlation (Figure 1 would be an example) and that all points are on (or very close to) the correlation line does not mean that the related molecules will lead to equally valid and simple additivity. It may happen that, for a few molecules, *both* properties show a similar deviation from an additivity. Such contributions would »cancel« each other in the correlation, and the point representing a molecule will be »in line« with others, yet its structure may involve additional contributions (*e.g.* close non-bonded interactions in overcrowded regions of the molecule). Similarly, if in such a property-property correlation a single molecule shows marked deviation for one property only, the corresponding point in the correlation diagram will be visibly displaced (see Figure 1 for illustration). Without further study it is, however, not possible to speculate which of the two properties has anomalous behavior. Of importance and interest in all such correlations is the *order* (along the correlation line) in which molecules occur, but this aspect of the correlations has hardly been considered. We will return to this later when discussing structure-property correlations.

STRUCTURE-STRUCTURE CORRELATIONS

Counterpart to *empirical* property-property correlation is *theoretical* structure-structure correlation, in which one set of structural parameters is compared and plotted against another set. If two models are equivalent, a simple equality of the type $y = x$ should result, but theoretical models frequently employ different approximations and the corresponding correlation is not usually such a simple line. The shape of the correlation, the dispersion of the points and other details can then be instructive in an analogous way in that such observations resulted in useful deductions when examining property-property correlations. As an illustration of structure-structure correlation (*i.e.* theoretical property *vs.* theoretical property), we show in Figure 5 a plot of Hückel molecular orbital (HMO) bond orders against valence-bond (VB) resonance-theoretical bond orders for a selection of benzenoid hydrocarbons. The HMO bond orders are taken from Coulson,[11] and Coulson and Streitwieser,[12] while the VB resonance-theoretical bond orders from Pauling *et al.*[13] and Herndon.[14]

We have selected this particular example since both concepts are well-known and not because these simple theoretical models are of great use today. Observe the scatter of points (corresponding to the CC bonds in selected benzenoid hydrocarbons) which indicate limited overlapping (conceptual and computational) of the Pauling bond orders and the Coulson bond orders. We know that, if the two theoretical models are further improved, the differences should reduce and eventually a respectable correlation may result. What Figure 5 confirms is that the two simple quantum chemical models disagree, at least in their prediction of bond lengths. It appears that both models are more or less equally limited in predicting the both lengths, but such an evaluation is beyond the scope of the present discussion. It is of interest to compare the HMO bond orders with the bond weights $(ij)^{-1/2}$, where i and j are valences of carbon atoms making up bond i–j in the bare carbon skeleton of a benzenoid hydrocarbon. The sum of all bond weights in a molecule gives the connectivity index χ:[15]

$$\chi = \sum_{\text{bonds}} (ij)^{-1/2} \tag{1}$$

Thus, the connectivity index is based on the differentiation of bond types. This was recognized as important already in 1947 by Hartmann.[16]

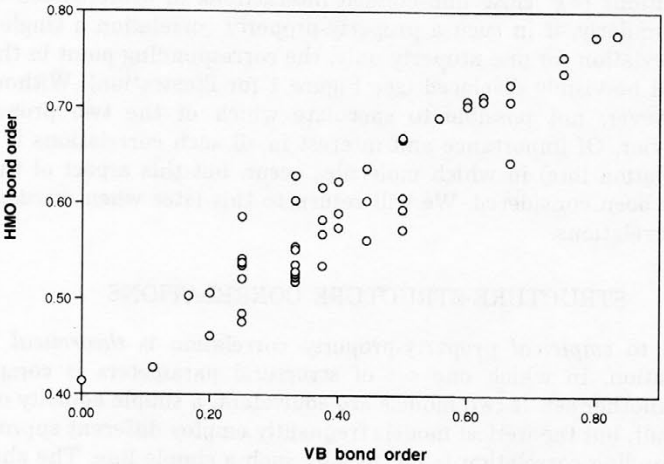

Fig. 5. A plot of the HMO bond orders *vs.* VB bond orders for selected benzenoid hydrocarbons.

The bond weights $(ij)^{-1/2}$ can also be viewed as primitive bond orders. Since the total π-electronic energy (E_π) in the normalized form of the HMO model ($\alpha = 0, \beta = 1$) can be expressed in terms of bond orders p_{rs}:

$$E_\pi = \mathrm{A} \sum_{\text{bonds}} p_{rs} + \mathrm{B}' \tag{2}$$

where $\mathrm{A}' = 2$ and $\mathrm{B}' = 0$, one can replace $\sum\limits_{\text{bonds}} p_{rs}$ by $\sum\limits_{\text{bonds}} (ij)^{-1/2}$ and express E_π in terms of the connectivity indices:

$$E_\pi = \mathrm{A} \sum_{\text{bonds}} (ij)^{-1/2} + \mathrm{B} \tag{3}$$

A plot of E_π against the connectivity index for 30 randomly selected benzenoid hydrocarbons is shown in Figure 6.

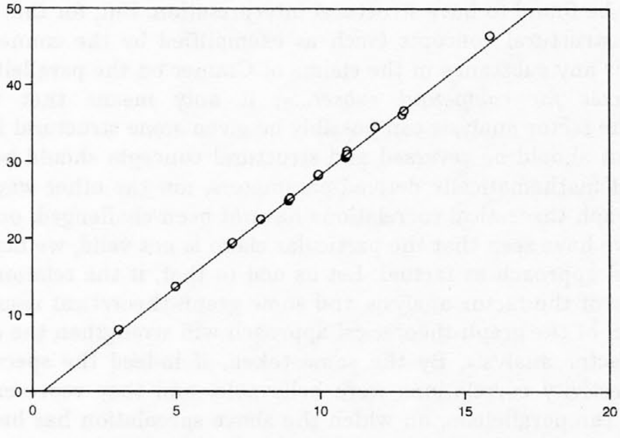

Fig. 6. A plot of the HMO π-electronic energies (E_π) *vs.* the connectivity indices (χ) for 30 randomly selected benzenoid hydrocarbons.

Surprisingly, we observe a much improved correlation. Why should suddenly such a simple *ad hoc* structural concept as the bond weight, devoid of any »quantum chemical« peerage, produce an impressive correlation with approximate quantum chemical quantity? This question may not be easy to answer. However, if we consider the HMO π-electronic energy as a theoretical »property« of molecules, the correlation in Figure 6 represents in fact a structure-property, not a structure-structure correlation. It allows interpretation of the HMO π-electronic energy in terms of bond types and connectivity indices. Ultimately, theoretical models and concepts are judged on how well they reproduce experimental quantities. Structure-structure correlations may illustrate, if not illuminate, similarities and differences between different models more clearly than tabular data or complex algebraic expressions.

STRUCTURE-PROPERTY CORRELATIONS

In order to proceed to structure-property correlations or even to the discussion of properties in terms of structure, one must have a *model* for the structure and such a model should be based on well-defined structural concepts. Hence, no discussion of an *empirical* approach devoid of structural components can produce an insight in structure-property correlations. In discussing the results of factor analysis (which is a purely empirical scheme), Cramer[8] speculates (citing no evidence) on a possible relation of his findings and available graph-theoretical correlations that were known for most of the considered properties, expressing the following opinion: »...*it seems reasonable to suppose that the molecular connectivity correlations are artifacts, perhaps representing alternative axes for compound subset*...« In view of what has been said, such an opinion is at best an incorrect surmise that perpetuates the confusion between *empirical* and *structural* approaches. If anything, it seems reasonable to suppose that in relating molecular connectivity to factor analysis it is conceivable that some results of the factor analysis can be *interpreted* in structural terms, *i.e.* some of the empirical parameters may be found to have structural interpretation. But, for this to be possible, one has to *use* structural concepts (such as exemplified by the connectivity index). Hence, if there is any substance in the claims of Cramer on the parallelism of the two »...*alternative axes for compound subset*...«, it only means that non-structural parameters of the factor analysis can possibly be given some structural interpretation. So, the argument should be reversed and structural concepts should be used for the interpretation of mathematically derived parameters, not the other way round. Since the *quality* of graph-theoretical correlations has not been challenged, only their interpretation, and we have seen that the particular claim is not valid, we may confirm the graph-theoretical approach as factual. Let us add to that, if the relation between the empirical results of the factor analysis and some graph-theoretical results is verified, the reaffirmation of the graph-theoretical approach will strengthen the significance of the results of factor analysis. By the same token, if indeed the speculation of the molecular connectivity correlations were believable, and they represented artifacts, then because of the parallelism, on which the above speculation has been based, this would mean that also BCDEFG axes used for the compound set are artifacts. The situation is, however, very simple: graph-theoretical correlations employ well-defined structural invariants and are reproducible, *e.g.*[17] The quality of such correlations is impressive[18] and this is probably the main cause for various speculations. How can a single (or a few simple) parameter(s) produce such high correlations? And then, in addition, a parameter (the connectivity index) is used that appears to be constructed in an *ad hoc* manner! Perhaps we should admit that the success of simple graph-theoretical correlations (those involving the connectivity indices,[19–21] but also other graph-theoretical indices[22,23]) may appear somewhat mysterious. We will now try to give some explanations and show that with proper background information available the »surprise« is not so great.

Many early approaches to bond additivities of thermodynamic properties used structural concepts without emphasizing them.[24] Such approaches are typically based on expressions involving from two to twenty parameters.[25] As time progressed, a call for higher accuracy of the additivities was justified. However, the desire for an ever-increasing capability of a model to reproduce data more accurately, may have at the same time caused neglect of a more careful analysis of simple models with very few parameters. The question to consider is *why* such simple models with a few parameters can perform so well – but the question has usually been overlooked.

The first simple successful model of this kind was due to Wiener,[26] who introduced only two parameters as variables in the study of isomeric variations of alkane thermodynamic properties. Wiener's parameters were: the total number of paths in a structure (denoted by W) and the total number of paths of length three (denote by p). In his own words:[26] »*The path number W is defined as the sum of the distances between any two carbon atoms in the molecule, in terms of carbon-carbon bonds.*«... »*The polarity number p is defined as the number of pairs of carbon atoms which are separated by three carbon-carbon bonds.*« The reader is reminded that *metrics* is defined for graphs, although graphs are mathematical objects. To the uninitiated in graph theory, the notion of »topological distance« may appear confusing. However, the concept of »distance« is more general than its common geometrical illustration. It requires the following conditions to be satisfied: (1) For any pair (x,y) of elements one associates a positive number D (usually called distance); (2) The direction of measurement of the distance is not important, *i.e.* $D(x,y) = D(y,x)$; (3) $D(x,x) = 0$; and (4) The so-called triangle condition: $D(x,y) \leq D(x,z) + D(z,y)$, signifying that the direct route is the shortest route between x and y. When »the number of intervening bonds« is taken as a measure of the »distance« between the atoms in a molecule, one speaks of »topological distance« (the concept introduced over a hundred years ago by Cayley[27]). It is easy to verify that the number of intervening bonds is a graph invariant[28] and that the above axioms of metrics are satisfied.

The Wiener model is truly a remarkable model, the significance of which Platt immediately recognized,[29] tried to justify and advertise, but apparently without much success. The work of Wiener and Platt has not been forgotten, it has been frequently cited during the past 46 years, but it did not have the impact it ought to have had in influencing subsequent research towards close examination of the virtues of such topological schemes. A new phase in structure-property correlation originates in the revival of interest in the chemical graph theory.[30–35] These more recent interests are centered on the design of a *single* parameter approach which would be based on some structural characteristics of molecules. The effort is to capture much of pertinent structural features in a single quantity. The first such approach is due Hosoya,[36] who designed an index (he called it the topological index, but it became known in the literature as the Z-index or the Hosoya index, *e.g.*[37]) as a sorting tool for classifying structures and immediately recognized that the same index is useful for the study of structure-property correlations.

The Wiener index, the Hosoya index, the connectivity index, *etc.* are generally referred to as topological indices, *e.g.*,[38] because they do not depend on details of molecular geometry, but only on molecular connectivity. More correctly, they should be referred to as *graph-theoretical indices*, but the term *topological index* is also acceptable.

The particular index of Hosoya, the Z-index, is based on the Z-polynomial (which is related to the acyclic polynomial[39]) of a molecular graph (molecule). It is defined as:[36]

$$Z = \sum_{k=0}^{[N/2]} p(G;\, k) \qquad (4)$$

where p(G,k) represents a selection of k mutually independent edges in a graph G, N/2 in the Gauss square brackets is the nearest integer not exceeding the real number in them and N is the total number of vertices in G. It should be noted that $p(G;0) = 1$ by definition and $p(G;1) =$ the number of edges in G.

Just as some veil of mystery accompanied Wiener's topological approach and the origin of Wiener's parameters W and p, so similarly in the case of Hosoya's topological index Z, one is perhaps uncomfortable with the astonishing success of such an *ad hoc* procedure. Apparently, the connectivity index[15] is yet another *ad hoc* topological index, which has been shown to lead to amazingly successful correlations of diverse molecular properties.[40] Before attempting to explain some of the reasons for achievements of these topological approaches, a comment pertaining to their *ad hoc* character is in place. It appears that, in some circles, exploring an area in an *ad hoc* manner is viewed as contemptuous. But, many important insights in science in general and in chemistry in particular followed some of such *ad hoc* considerations. Bohr's model of the atom is an *ad hoc* construction and so is Schrödinger's »derivation« of the wave equation based on analogy with waves in continuous media. It was only subsequently found that such *ad hoc* procedures have a much wider applicability. In fact, by virtue of its not being related to previously known schemes, and *ad hoc* method is likely to bring *novel* insights and relationships to light, and may prove to be a more desirable scheme in areas where general principles are not known. The situation is perhaps well illustrated by a quote from Van Vleck's Nobel Prize lecture[41] concernig the outstanding early attempts to understand magnetism at the atomic level: »...*To account for paramagnetism, Langevin in 1905 assumed in a purely* ad hoc *fashion that an atomic or molecular magnet carried a permanent moment* μ*, whose spatial distribution was determined by the Boltzmann factor. It seems today almost incredible that this elegantly simple idea had not occurred earlier to some other physicist inasmuch as Boltzmann had developed his celebrated statistics over a quarter of a century earlier...*« In fact, one could make almost the same remark concerning the Z-index, the connectivity index χ, acyclic polynomials, conjugated circuits[42] or isomerisation graphs,[43] that it seems almost incredible today that these elegantly simple ideas did not occur earlier to some other chemists inasmuch as Hückel had developed his celebrated model,[44] Pauling advanced the concepts of quantum mechanics for chemistry[45] and Polyá formulated his counting theorem,[46] all more than half a century earlier!

The connectivity index is a bond additive quantity, which makes it suitable for bond additivities. It is based on a scheme of preweighted contributions. The weighting factors, which are fixed, have been determined in such a way that the resulting indices for isomers produce an *ordering* of structures that has been observed to hold for many thermodynamical properties. Formally, the conditions for a relative ordering of structures lead to a system of inequalities and the »*ad hoc*« recipe for the construction of bond contributions presents a particular solution to the system of inequalities. The whole approach is clearly based on a number of assumptions, including selection of a particular form for the solution of inequalities (which may generally have many alternative solutions), and hence the approach should be judged by its performance. As already mentioned, the connectivity index passed the test exceptionally well. The index has been generalized[47] to include larger fragments as contributing elements, extended to incorporate heteroatoms[48] and applied to a large number of structure-property and structure-activity problems. This includes the topic of drug design,[49] the use of the connectivity index as an auxiliary structural descriptor in computer-assisted studies of chemical structure and biological function[50] and in the classification of drugs by discriminant analysis.[51] Review of many such graph-theoretical correlations can be found in, for example, books by Kier and Hall[40,52] and a number of review articles, *e.g.*[53] Perhaps, the success of the connectivity index is in part responsible for the continuing interest in topological indices and their proliferation. At the same time, many have been surprised by the success of an index that is apparently based on some arbitrary concepts, so that a fuller justification of its performance appears desirable. We will address both of these aspects with some remarks on the evaluation of alternatives and will try to develop some guiding rules that topological indices ought to satisfy in order to qualify as useful and novel, not simply being a reformulation of the known. However, at this point we emphasize that the best approach to calculating physico-chemical properties depends on the property and that at present *no approach* is adequate to calculate *all* the physical and chemical properties of a molecule.

RULES FOR THE SELECTION OF TOPOLOGICAL INDICES

Some structural components, such as atoms and bonds, are so natural a choice and so clear that there is hardly any dispute on their role and significance in molecular additivities. On the other hand, it is far from clear why the total number of paths, the parameter W of Wiener, the sum of the coefficients of acyclic polynomial, the parameter Z of Hosoya or factors $(ij)^{-1/2}$ for bond types (i,j) in the case of the connectivity index, should be important or even relevant to molecular additivities. We are going to clarify the situation. Clearly, atoms and bonds *are not enough* when one is interested in details of isomeric variations of numerous molecular properties and going beyond atoms and bonds is not unique. First, difficulties arise here concerning the number of such additional parameters and their choice. Whilst many have recognized the particular choice of parameters W and p in the Wiener approach as arbitrary, the prescription for Z-index as arbitrary or the construction of the connectivity index as arbitrary, few realize that already the selection of atoms and bonds, and then, what appears as a »natural« extension, selection of next-nearest neighbors *etc.* is equally arbitrary. Selection of *any* component is an arbitrary act and should be viewed as analogous to the selection of a coordinate system for the computational problem considered. Frequently, very different coordinates serve well specific problems, so in analogy one may expect a widely different collection of graph invariants to suit specific applications. The coordinates, like structural invariants, serve the purpose of comparing observable quantities, they do not constitute the results but help express the results. In analogy to a choice of coordinates, graph invariants can, therefore, simplify the resulting expressions or reduce the amount of computations involved.

The first step in devising structure-property correlations should concern the *number* of parameters to be used. Once the correlation or an additivity is established, the adopted parameters are assigned numerical values and the large number of parameters generally require a large sample of structures if one wishes to maintain some prescribed statistical significance for the parameters.[54] Occam's dictum: »*Essentia non sunt multiplicanda praeter necessitatem*« should be here interpreted as »parameters should not be multiplied without necessity« or one should use as few parameters as possible. The problem is how to arrive at a useful qualification of the »necessity«. One way of arriving at some practical *modus operandi* is to require that the search for the structural correlations be conducted in a manner that *reports* the best correlation when a single parameter is used, followed by the best correlation based on two parameters, and so forth. In this way, for a collection of structures and a selected property, a meaningful comparison between different models is possible. If a claim is made on superiority of one model over another, all that one has to ensure is that the compounds selected are the same *and* that no deliberate omissions of any structures belonging to the class considered are made. The latter request is important because two models may differ significantly precisely in the few selected structures. By accepting such a course of action, one would be able to speak of the best 1-parameter correlation, best 2-parameter correlation, *etc.* For many well-established correlations we do not know today what would be the best 1-parameter approach, the best 2-parameter approach, *etc.* For instance, does the use of the connectivity index in the structure-property-activity relationships lead to the best 1-parameter scheme?, is the Wiener's use of W and p the best 2-parameter scheme? *etc.* Very few such studies have appeared, a recent one was concerned with the comparison of several 4-parameter bond energy schemes.[55] Models that give the same or similar results for the same collection of compounds and the same number of parameters can then be considered equivalent, even if the relationship between the two sets of invariants is not apparent. Such models can be further judged by extending the pool of properties and by increasing the pool of structures, if appropriate.

One anticipates that for a *different* property, a different structural scheme may be preferred. Availability of a few successful indices does not automatically preclude the need for other indices. But, new indices have to show *novel* features in order to represent a valuable addition. Hall[56] appears to be the first to consider a condition for topological indices in order to discriminate among many available alternatives. As he indicated, there is ample evidence that many of numerous thermodynamical properties depend on the role of the next-nearest neighbors. Accordingly, indices which do not incorporate some dependence on the next-nearest neighbors are bound to be deficient for describing such properties, although they may have other uses outside the considered field. This approach seems prudent and worth expanding. In other words, indices which do not take into account influences of the next-nearest neighbors will not qualify as best 1-parameters and appear not to present a good start for building more elaborate additivity schemes.

RULES ON THE ORDERING OF STRUCTURES

The next requirement that can be imposed is that the scheme (*e.g.*, a single parameter model) reproduces the same *order* for structures as the order determined by the property considered. Small numerical differences need not be recognized and such structures can be grouped. This requirement appears to be a trivial one, yet in applications a number of indices are found that do not satisfy it. Consider various thermodynamical properties of alkanes (boiling points, heats of vaporization, and so on) for *n*-alkanes, 2-methyl and 3-methyl derivatives. One finds that the relative order for the three isomers is as follows: *n*-alkane, 3-methyl, and 2-methyl derivative; 3-methyl substituted isomers are invariably always more similar (numerically) to the value of *n*-alkane. A request for the ordering of structures requires the structural parameters for the same isomers to define the *same* order. One finds that the Hosoya Z-index and the connectivity index,[57] which provide a better discrimination among isomers, do not predict correct relative

TABLE I

Isospectral nonanes (that is, nonanes having the same eigenvalue spectra) and their selected properties: (a) boiling points (in °C); (b) heats of vaporization (kcal/mol); (c) indices of refraction; (d) liquid densities (g/cm^3); (e) constants A in the Antoine equation and (f) the Z-indices. The last two rows give the range of values for the above properties in all nonanes.

Isospectral pair	Properties					
	(a)	(b)	(c)	(d)	(e)	(f)
2,3-dimethylheptane (**1**)	140.5	132.21	1.409	0.726	6.889	44
4-ethyl-2-methylhexane (**2**)	133.8	148.95	1.407	0.723	6.869	44
2,3,5-trimethylhexane (**3**)	131.3	145.60	1.406	0.722	6.856	37
2,2-dimethylheptane (**4**))	132.7	145.60	1.402	0.711	6.856	37
2,3,4-trimethylhexane (**5**)	139.0	149.37	1.414	0.739	6.877	41
3,3-dimethylheptane (**6**)	137.2	148.53	1.409	0.725	6.867	41
2,4-dimethyl-3-ethylpentane (**7**)	136.7	148.11	1.414	0.738	6.866	39
4,4-dimethylheptane (**8**)	135.2	148.11	1.408	0.725	6.856	39
2,2-dimethyl-3-ethylpentane (**9**)	133.8	145.60	1.412	0.735	6.855	36
2,3,3-trimethyhexane (**10**)	137.7	146.44	1.414	0.738	6.866	36
Maximal value	150.8	158.16	1.424	0.757	6.912	55
Minimal value	122.3	137.24	1.400	0.707	6.811	24

ordering for the selected properties. We ought to stress the *selected*, because there are other properties (including thermodynamical properties) which dictate different ordering of isomers. For example, the octane numbers [58] and mean squared radius[59] of alkanes follow the order: *n*-alkane, 2-methyl, and 3-methyl derivatives, and for these properties information-theoretic indices[57] and centric indices[58] are clearly more suited than Z or χ.

A new problem arises with the use of more than one parameter: *comparison* of structures needs to be defined. A situation may arise that the structures are not comparable. This problem has been considered in physics[60] and chemistry[61] and has been discussed[62] on the basis of the early rules proposed by Muirhead at the beginning of this century.[63–65] Ordering of sequences (which may represent structures) may be based on alternative ordering rules. One may view the parameters as coordinates of a structure in the corresponding »structure space«. An illustration is given by the use of path numbers p_2 and p_3, which produces a two-dimensional ordering of isomers and results in a grid or periodic table for isomers.[66] Finally, we may mention yet another rule for the ordering of structures, based on a comparison of atomic environments, which »induces« a partial order among structures.[67] The choice of the ordering rule will depend on the problem and the property. If a regularity is found in a property, one concludes that the particular rule adequately reflects the ordering of structures. If the same regularity is found for selected structural parameters, one may be on a valid track to establishing a structure-property correlation. *Ordering* as an operation was overlooked in the past although quantitative structure-property and structure-activity studies should consider the ordering of structures as one of the first steps in their efforts to derive a correlation or an additivity. Let us mention several advantages of the ordering of structures, which will illustrate the importance of this step in structure-property-activity studies:

(1) The validity of a model can be tested on isocodal structures. Isocodal structures are those having all the selected parameters equal and, consequently, predict the same (or similar) property. If such isocodal structures shows appreciable variation, the model (the selected invariants) cannot provide an adequate basis for structure-property correlation;

(2) Deviations of some points from the correlation can be attributed to the *incompatibility* of structures. Hence, an apparent limited correlation may have a valid structural cause, and can even be anticipated;[68]

(3) Ordering of structures may lead to unsuspected correlations;[69]

(4) »Unrelated« data can be recognized as having a common trait.

To further illustrate some of the mentioned points, let us consider Hosoya's Z index. We have the same index for isospectral molecules[31,70] (see Table I) and, therefore, predict the same magnitudes in a structure-property correlation using Z index. (Isospectral structures are non-isomorphic structures that have identical eigenvalue spectra.[71]) In Figure 7, we have illustrated the carbon skeletons for the isospectral non-

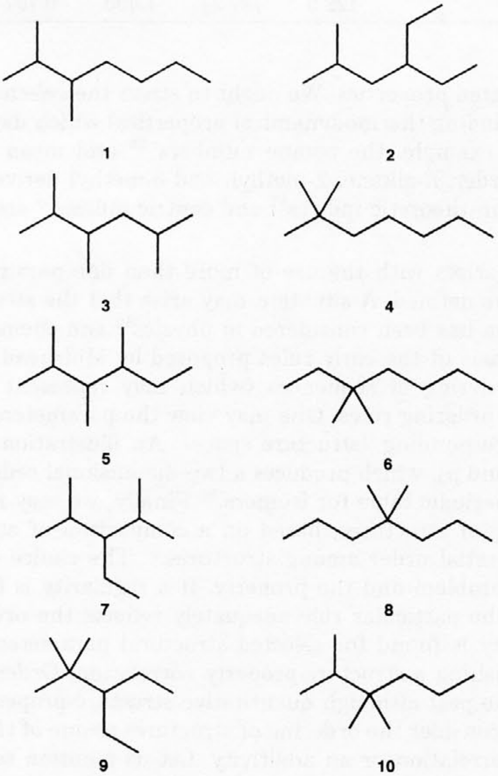

Fig. 7. Carbon skeletons of isospectral nonanes.

anes C_9H_{20} considered in Table I, from which one sees that the paired structures indeed have limited apparent similarity.

In a number of cases, already vertex valencies for two structures are different, and even when the same, the molecules differ in bond types. Hence, there are no »visible« reasons why such molecules should show similar properties, but they do. In a few instances boiling points show somewhat larger deviations, which are still not excessive in view of the fact that boiling points cover a range of 28.5 °C in nonanes. Moreover, coincidences are here more significant than some disagreements because there are other factors, such as nonbonded close contacts, which are not taken into account and which will make significant differences between isospectral pairs. The fact is, that, for example, 2,3,5-trimethylhexane and 2,2-dimethylheptane have very similar boiling points, indices of diffraction and liquid densities, and identical heats of vaporization and the Antoine constants A. This is either a pure coincidence or is possibly due to some structural cause. Our conjecture is that these are not pure »coincidences«. The Hosoya index Z suggests that this is the case. Overall, we may say that Table I displays a fairly good evidence in support of the notion that isospectral structures may be expected to show similar thermodynamical properties. One has to be mindful that we are employing a *single* structural parameter and that, therefore, there are limits to accuracy for other (non-isospectral) structures as well and the quality of the agreement for all (isospectral and non-isospectral) structures is expected to be approximately the same.

We conclude that the concept of isospectral graphs, which has received some attention in the literature,[72] is relevant in chemistry, in contrast to the opinion derived from the consideration of a single property of some of such molecules. Upon comparing photoelectron spectra of 1,4-divynilbenzene and 2-phenylbutadiene (which form an isospectral pair), Heilbronner and Jones[73] expressed the opinion that »*...the graph theoretical statement that two molecules are "isospectral" is of no relevance for their physical and chemical behaviour...*« because they found the actual spectra to be different. The expectation that isospectral molecules may have similar spectra follows from the interpretation of the graph adjacency matrix as the interaction matrix (in the Hückel MO method). But the Hückel MO has been known to be deficient for discussion of π-electron spectra for some 40 years. In fact, the recognition of this deficiency has stimulated an early interest in the development of more sophisticated approximate MO theories, exemplified by the Pariser-Parr and Pople methods.[74,75] Thus, an effort to investigate photoelectron spectra of isospectral conjugated systems boils down to verification of whether a known deficient scheme is equally deficient in different molecules! This may still be a valid task, but perhaps peripheral to current theoretical interests. In addition, however, the conclusion is incorrect and could have been avoided if the authors had taken molecular graphs and associated concepts in a broader context, not limiting (as, unfortunately it has frequently been the case among quantum chemists) such considerations to π-electron conjugated systems only. Our considerations here clearly apply to isospectral non-conjugated systems and, when discussing the concept if isospectral graphs, already in the introductory part of one of the papers[76] we indicated a wider socpe for the concept when stating: »*...Generally, the topological eigenvalues and eigenvectors have distinctive significance and should reflect the* inherent properties of molecular skeletons *as such. Their study is of interest then in investigations of molecular properties which critically depend on molecular size, shape and conformation, degree of branching and its mode, etc. This will enable one to trace some of the molecular properties to a topological origin.*« The statement requires no revision, perhaps attention!

As another illustration of the use of the isocodal test, consider the isocodal structures in Table II having the same values for p_2 and p_3.

We have selected 7 properties and have indicated the range of values for each property (the maximal and the minimal values found for other isomers), which provide some indication of the degree of agreement among the isocodal structures. Additional comparisons of this kind can be made by considering, for example, not octanes[77] but nonanes,[78] which in view of a larger number of isocodal cases make accidental coincidences very unlikely. Table II is indicative of the potential use of the structural parameters p_2, p_3 for discussion of molecular properties.

TABLE II

Isocodal octanes (that is, octanes having the same p_2,p_3) and their selected properties (a) critical densities (in g/cm^3); (b) critical pressures (in atm); (c) specific dispersions (in cm^3/g); (d) heats of combustion (in kcal/mol); (e) critical temperatures (in °C); (f) critical volumes (in cm^3/mol); (g) mean radiuses (in Å) and (h) connectivity indices. The last two rows give the range of values for the above properties in all octanes

Isocodal structures	Properties							
	(a)	(b)	(c)	(d)	(c)	(f)	(g)	(h)
3-ethyl-2-methylpentane	0.254	27.4	96.2	1222.11	295	0.450	1.521	3.718
3,4-dimethylhexane	0.253	27.4	96.7	1221.68	298	0.452	1.552	3.718
3-methylheptane	0.239	26.6	97.6	1221.76	292	0.478	1.798	3.808
4-methylheptane	0.240	25.6	97.6	1221.89	290	0.476	1.767	3.808
Maximal value	0.263	29.0	106.6	1222.70	305	0.489	2.044	3.914
Minimal value	0.234	24.5	95.9	1218.59	271	0.433	1.250	3.250

A HIERARCHY OF STRUCTURAL INVARIANTS

Selection of graph (or structural) invariants is left to individual researchers, just like the choice of coordinates or basis functions in quantum chemical computations. Hence, it is not surprising to see many different choices. One may be interested in developing a »system« by selecting invariants that would suffice for characterization of any structure in a useful manner, but, while such an effort appears attractive, it may be more ambitious than necessary. Many molecules have the same or similar properties and if one is interested in structure-property correlations, one does not need to have different codes for molecules, since some will have the same properties and, hence, can be assigned the same descriptors. All one should require is to capture the most relevant structural features responsible for the variation in magnitude for the property considered. A complete set of graph invariants is not yet available which could serve as a reference for converting different structural parameters to the same reference scale. It is questionable whether such a complete set is possible! We suggest here a more modest option: a selection of *reference* invariants that can be classified in the order of increasing generality. Generality is here to be interpreted as the capacity for preserving structural information. Although such *reference* invariants do not constitute a complete system (which would consequently lead to unique structural names), they can provide a useful basis for comparison of different models. We will confine our discussion to molecular graphs rather than molecular structures. Then, the following appear as useful *reference* invariants, given in a hierarchical order of an increasing level of structural information content:

(1) the count of neighbors at different distances;
(2) the count of paths (or self-avoiding walks) of different length;
(3) the count of self-returning walks;
(4) the count of random walks.

In acyclic structures, the count of neighbors and the count of paths are equivalent since any two vertices are connected by a single path. In cyclic structures the count of neighbors becomes the count of the shortest paths, which are entries in the distance matrix. Algorithms for the above counts have been outlined and computer programs are available,[79] while the count of self-returning walks and random walks is relatively simple.[80] It amounts to evaluation of the powers of the adjacency matrix **A**.[81] In Figure 8, we have illustrated all the self-returning walks of length six and have also depicted their alternative representation as subgraphs.

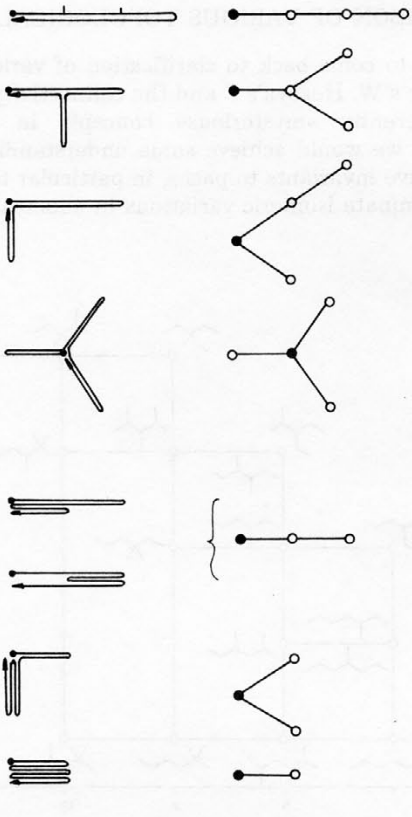

Fig. 8. Self-returning walks of length six and the corresponding subgraphs representing them. The cases related to acyclic graphs only have been considered. The starting (and ending) vertex is indicated as a black dot.

It would be of interest to know the number of all subgraphs once matrices $\mathbf{A}^m$ have been evaluated, the problem which has not yet been considered in the literature. It appears that information on smaller subgraphs of this kind, corresponding to self-returning walks of shorter distance (radius), can be related to the degree of a vertex,[82] but the topic has to be further studied. Self-returning walks have shown some unusual properties,[83,84] and deserve more attention. The relation between paths and walks has been considered, and a simple connection is found in graps for paths that are shorter than the periphery of the smallest cycle.[85] The hierarchical ordering of invariants is a new concept, but it has already been found useful in an attempt to extend the concept of graph center (well-defined previously only for acyclic structures by Jordan[86]) to polycyclic graphs.[87,88] In the next section, we will use one of the reference invariants (the concept of paths) and will show how such invariants can be used to compare different topological indices. As a result, we will also see the reasons why such simple 1-parameter approaches to structure-property correlations work reasonably well and, we hope thus to lift the veil of mystery that accompanied these important analytical-theoretical avenues to the study of molecules.

COMPARISON OF VARIOUS TOPOLOGICAL INDICES

We would now like to come back to clarification of various graph-theoretical invariants, such as Wiener's W, Hosoya's Z and the connectivity index χ. If we could express the above apparently »mysterious« concepts in terms of other more »transparent« concepts, we would achieve some understanding of their content. We will try to relate the above invariants to paths, in particular to paths p_2 and p_3, which we already found to dominate isomeric variations in alkanes.[77,78,89,90] In Figure 9, we

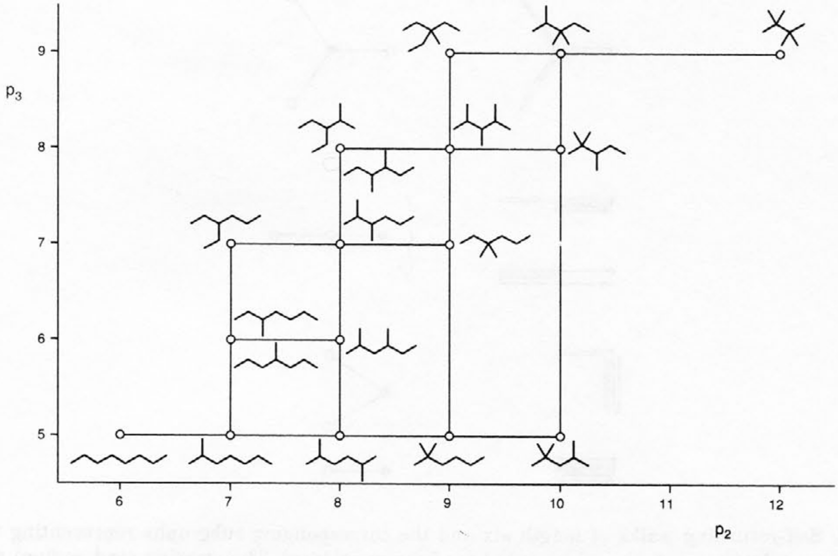

Fig. 9. The (p_2,p_3) grid for octanes isomers in which isomers are placed at the corresponding (p_2,p_3) values.

show a grid with points representing various octane isomers. For each isomer, one counts the fragments C–C–C and C–C–C–C which give p_2 and p_3, respectively, and determine the coordinates (p_2,p_3) for the isomer.

In the case of coordinates (7,6) and (8,8), we find two isomers to belong to the same location (as already indicated in Table II). The fundamental hypothesis is that structural *and* molecular properties vary with p_2 and p_3 rather »smoothly«, in harmony with the well-known Aristotle's dogma: »*Natura non facit saltus*«, *i.e.*, the changes are gradual (the principle of graduality).[91] The validity of the approach has already been tested on many molecular properties of alkanes, *e.g.*[92] Here, we want to show that the same approach is valid for *structural indices*. This will then provide an answer why simple topological approaches work so well. They are dependent on the same fundamental (reference) invariants in the same way as properties.

Figure 10 illustrates the (p_2,p_3) grid in which each isomer is replaced by the numerical value of (a) Wiener's W number; (b) Hosoya's Z index; (c) the connectivity index χ; and (d) the information-theoretic index I_D of Bonchev and Trinajstić.[57,93] In brackets, the signs of change along p_2 and p_3 are shown which classify Z and χ as one class, and W and I_D as another class. A meaningful comparison is possible only between indices of the same class because they can be employed for the same collection of molecular properties, which have to be of the same class.

Observe that four of the mentioned indices vary relatively smoothly along p_2 and p_3. There may be a small oscillation in the differences between adjacent values, but

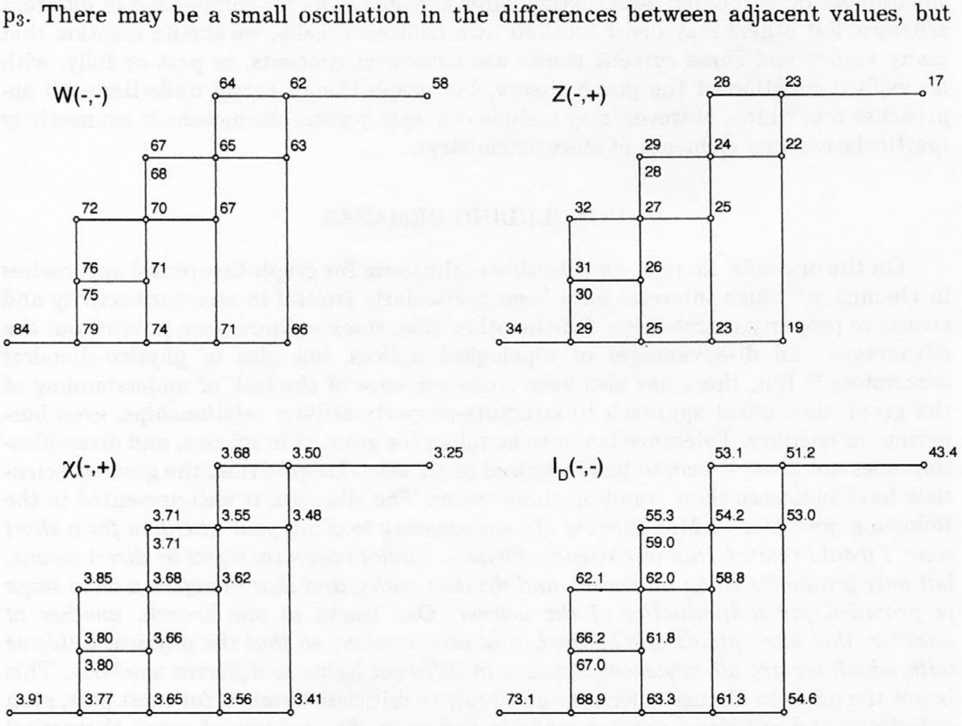

Fig. 10. Isomeric variations of selected topological indices with p_2 and p_3 path numbers for octanes: (a) Wiener's index W; (b) Hosoya's Z-index; (c) the connectivity index χ and (d) the information-theoretic index I_D.

the trends generally appear to be well-reproduced. We conclude that all the four indices considered qualify for discussion of molecular properties, displaying desirable *regular* changes with p_2 and p_3, although they will not all be equally suitable for the same properties.

Selection of graph invariants may be influenced by previous work and considerations, or based on what appears to some as »natural« or can be unrelated to earlier ideas.[94–96] Conceptually, path numbers p_2 and p_3 are very simple and may, therefore, provide the basis for interpretation of other successful approaches. They certainly have a more transparent interpretation than the connectivity index χ, the topological index Z, and Wiener parameter W, which can all be viewed as *derived*. Since the mentioned indices show a regular variation with p_2, p_3, we can expect them to provide useful parameters for structure-property correlations. Occasionally, the question on the fundamental or non-fundamental character of a particular parameter is raised. Should formal carbon atom valencies (in hydrogen-suppressed graphs),[97] *i.e.*, the concept of primary, secondary, tertiary and quaternary carbon atoms, be considered more fundamental than path numbers p_k? Should not the count of conformational isomers (such as rotational isomers) be considered a fundamental factor? Various additivity schemes used various structural parameters considering them fundamental, and it still appears to be a matter of personal preference what one calls fundamental. We would like to call p_2 and p_3 fundamental (in addition to p_0 – the number of atoms and p_1 – the number of CC bonds) because they offer a simple *basis* for comparison of different schemes, but others may insist on their own choices. Finally, we should mention that many earlier and some current works use structural concepts, in part or fully, with no explicit mention of the graph theory, but graph theory partly underlies such approaches too, which, however, may include concepts beyond the molecular connectivity (particularly some elements of stereochemistry).

CONCLUDING REMARKS

On the one side, there is considerable enthusiasm for graph-theoretical approaches in chemistry.[98] Such interests have been particularly fruitful in structure-activity and structure-property correlations. On the other side, some authors have pointed out the advantages and disadvantages of topological indices, but also of physico-chemical descriptors.[99] But, there has also been some evidence of the lack of understanding of the graph-theoretical approach to structure-property-activity relationships, even bordering on hostility. Tolerance is not to be taken for granted in science, and diversification does not always seem to be recognized as an asset. Despite that, the greatest scientists have indicated their stand on these issues. The situation is well-presented in the following quotation of Max Planck: »*In endeavoring to claim your attention for a short time, I would remark that our science, Physics, cannot attain its object by direct means, but only gradually along numerous and devious paths, and that therefore a wide scope is provided for individuality of the worker. One works at one branch, another at another, this one applies one method, that one another, so that the physical universe with which we are all concerned appears in different lights to different workers*«. This is not the place to discuss objections and reply to criticism because, for most part, such objections and criticisms are not made in the open. The volume of graph-theoretical literature in chemistry has increased visibly during the last twenty years and one expects that critics have been given *ample* material on which they can focus their disapproval and engage in constructive dialogue. As they appear unwilling to air their reservations publicly, yet hinder the development of a theoretical branch by expressing opinions on a field to which they have made no contribution, it appears that it may be in order to solicit public support in an effort to clarify the role of the graph theory in chemistry. Traditionally, novel approaches (*e.g.*, quantum mechanics) or new applications of neglected disciplines (*e.g.*, the group theory in 1930–50 particularly in chemistry) have had difficulties before being accepted. At one time or another, many presently considered essential and fundamental approaches have experienced such initial receptions. Not long ago, perhaps, the pattern recognition as a tool in chemistry was in a similar position but it succeeded in enlisting distinguished support.[100] Such support may encourage newcomers and attract new talents, although its beneficial impact on those in the field already should not be belittled. Great scientists, like Planck, have appreciated such support, as evidenced from the quote: »*...As an offset against much disappointment I derive much satisfaction from the fact that Ludwig Boltzmann, in a letter acknowledging my paper, gave me to understand that he was interested in, and fundamentally in agreement with, my ideas.*« In concluding, we would like to make a similar remark. As an offset against much disappointment, we derive much satisfaction from the fact that Coulson, in a letter acknowledging a paper of one of us, gave us to understand that he was interested in the forthcoming graph-theoretical revival:[101] »*...I was particularly interested in your graph theory paper. This topic seems to be rather popular at the moment...*« This was in 1973, and perhaps reflects the general optimism that radiated around the late Professor Coulson, but it may take a few more years before such optimism is fully appreciated.

We started the presentation with a quote of Max Planck, borrowed a few more for appropriate parts of the text, and would like to end with yet another extract from Max Planck's Survey of Physical Theory:[2] »*It would certainly be a serious illusion on my part if I hoped that may remarks have carried general conviction, or even that they have been generally understood, and I shall very anxiously leave it with you. Surely much more will be thought and written concerning these questions, for theorists are numerous and paper is patient.*«

Acknowledgments. – This work was supported in part by the Ministry of Science and Technology of the Republic of Croatia through grants No. 1–07–159 and 1–07–185.

REFERENCES

1. E. Cohen, *Jacobus Henricus van't Hoff — Sein Leben und Wirken,* Akademisches Verlagsgesselschaft, Leipzig, 1912, p. 54.
2. K. Fajans, *Ber. Deut. Chem. Ges.* (1920) 643.
3. M. Planck, *Survey of Physical Theory*, Dover, New York, 1960.
4. E. A. Smolenskii, *Russ. J. Chem. Soc.* **38** (1964) 700.
5. M. Gordon and J. W. Kennedy, *J. Chem. Soc. Faraday Trans II* **69** (1973) 484.
6. W. R. Dillon and M. Goldstein, *Multivariate Analysis — Methods and Applications*, Wiley, New York, 1984.
7. A. L. Edwards, *Multiple Regression and Analysis of Variance and Covariance*, Freeman, San Francisco, 1979.
8. R. D. Cramer, III, *J. Amer. Chem. Soc.* **102** (1980) 1837.
9. E. R. Malinowski, *Factor Analysis in Chemistry,* Wiley-Interscience, New York, 1991.
10. W. J. Dunn, III, S. Wold, and Y. C. Martin, *J. Med. Chem.* **21** (1978) 922.
11. C. A. Coulson, *Proc. Roy. Soc. (London)* **A169** (1939) 413.

12. C. A. Coulson and A. Streitwieser, Jr., *Dictionary of π-Electron Calculations,* Freeman, San Francisco, 1965.
13. L. Pauling, L. O. Brockway, and J. Y. Beach, *J. Amer. Chem. Soc.* **57** (1935) 2705; L. Pauling, *Acta Cryst.* **B36** (1980) 1898.
14. W. C. Herndon, *J. Amer. Chem. Soc.* **96** (1974) 7605.
15. M. Randić, *ibid.* **97** (1975) 6609.
16. H. Hartmann, *Z. Naturforsch.* **2a** (1947) 259.
17. Z. Mihalić and N. Trinajstić, *J. Chem. Educ.* **69** (1992) 701.
18. D. E. Needham, I-C. Wei, and P. G. Seybold, *J. Amer. Chem. Soc.* **110** (1988) 4186.
19. A. Sabljić and N. Trinajstić, *Acta Pharm.* **31** (1981) 189; N. Trinajstić, M. Randić, and D. J. Klein, *ibid.* **36** (1986) 267.
20. D. H. Rouvray, *ibid.* **36** (1986) 239; M. I. Stankevich, I. V. Stankevich, and N. S. Zefirov, *Russ, Chem. Rev.* **57** (1988) 191.
21. A. Sabljić and H. Güsten, *Chemosphere* **19** (1989) 1503; A. Sabljić, R. Lara, and W. Ernest, *ibid.* **19** (1989) 1665; A. Sabljić, *Environ. Health Perspect.* **83** (1989) 179; A. Sabljić, J. Güsten, J. Schönherr, and M. Riederer, *Environ. Sci. Technol.* **24** (1990) 1321; S. Sekušak and A. Sabljić, *J. Math. Chem.* **11** (1992) 271; *J. Chromatogr.* **628** (1993) 69.
22. P. G. Seybold, M. A. May, and M. L. Gargas, *J. Chem. Educ.* **64** (1987) 575; I. Lukovits, *J. Chem. Soc. Perkin Trans. II* **84** (1988) 1667; *Quant. Struct. Act. Relat.* **9** (1990) 227; *Reports Mol. Theory* **1** (1990) 127; M. Medić-Šarić, S. Nikolić, and J. Matijević-Sosa, *Acta Pharm.* **42** (1992) 139; S. Nikolić, M. Medić-Šarić, and J. Matijević-Sosa, *Croat. Chem. Acta* **66** (1993) 151.
23. P. A. Filip, T.-S. Balaban, and A. T. Balaban, *J. Math. Chem.* **1** (1987) 61; A. T. Balaban and V. Feroiu, *Reports Mol. Theory* **1** (1990) 133; K. Osmialowski and R. Kaliszan, *Quant. Struct. — Act. Relat.* **10** (1991) 125; T.-S. Balaban, P. A. Filip, and O. Ivanciuc, *J. Math. Chem.* **11** (1992) 79; A. T. Balaban, *J. Chem. Inf. Comput. Sci.* **32** (1992) 23; O. Ivanciuc, T.-S. Balaban and A. T. Balaban, *J. Math. Chem.* **12** (1993) 309; *ibid.* **14** (1993) 21; A. Voelkel, T. Kopczynski, and A. Gawalek, *J. Mol. Struct. (Theochem)* **279** (1993) 217; T. Kopczynski and A. Voelkel, *ibid.* **279** (1993) 217.
24. V. M. Tatevskii, *Khimicheskoe stroenye uglevodorov i zàkonomernosti v ikh fiziko-khimicheskikh svoistv,* Nauka, Moscow, 1953; V. M. Tatevskii, V. A. Benderskii, and S. S. Yarovoi, *Rules and Methods for Calculating the Physico-Chemical Properties of Paraffinic Hydrocarbons*, Pergamon, New York, 1961.
25. S. W. Benson, F. R. Cruickshank, D. M. Golden, G. R. Haugen, H. E. O'Neal, A. S. Rodgers, R. Shaw, and R. Walsh, *Chem. Rev.* **69** (1969) 279.
26. H. Wiener, *J. Amer. Chem. Soc.* **69** (1947) 17; *ibid.* **69** (1947) 2636; *J. Chem. Phys.* **15** (1947) 766; *J. Phys. Chem.* **52** (1948) 425.
27. A. Cayley, *Cambridge Math. J.* **2** (1841) 267.
28. P. J. Hansen and P. C. Jurs, *J. Chem. Educ.* **65** (1988) 574.
29. J. R. Platt, *J. Chem. Phys.* **15** (1947) 419; *J. Phys. Chem.* **56** (1952) 328.
30. A. T. Balaban (Ed.), *Chemical Applications of Graph Theory*, Academic Press, London, 1976.
31. N. Trinajstić, *Chemical Graph Theory*, CRC Press, Boca Raton, FL. 1983, Vols. I and II.
32. R. B. King (Ed.), *Chemical Applications of Topology and Graph Theory,* Elsevier, Amsterdam, 1983.
33. I. Gutman and O. E. Polansky, *Mathematical Concepts in Organic Chemistry,* Springer-Verlag, Berlin, 1986.
34. R. B. King and D. H. Rouvray (Eds.), *Graph Theory and Topology in Chemistry,* Elsevier, Amsterdam, 1987.
35. N. Trinajstić, *Chemical Graph Theory,* CRC Press, Boca Raton, FL. 1992. 2nd revised edition.
36. H. Hosoya, *Bull. Chem. Soc. Japan* **44** (1971) 2323; H. Hosoya, K. Kawasaki, and K. Mizutani, *ibid.* **45** (1972) 3415.
37. D. Bonchev, *Information Theoretic Indices for Characterization of Chemical Structures,* Research Studies Press, Chichester, 1983, p. 70.
38. D. H. Rouvray, *Amer. Scientist* **61**(1973) 729.
39. I. Gutman, M. Milun, and N. Trinajstić, *Math. Chem. (Mülheim/Ruhr)* **1** (1975) 171; *J. Amer. Chem. Soc.* **99** (1977) 1692.

40. L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure-Activity Analysis,* Wiley, New York, 1986.
41. J. H. Van Vleck, *Science* **201** (1978) 113.
42. M. Randić, *Chem. Phys. Lett.* **38** (1976) 68.
43. A. T. Balaban, D. Farcasiu, and R. Banica, *Rev. Roumaine Chim.* **11** (1966) 1205.
44. E. Hückel, *Z. Physik* **70** (1931) 204; *ibid.* **72** (1932) 310; *Ibid.* **76** (1932) 628; *ibid.* **83** (1933) 632.
45. L. Pauling, *The Nature of the Chemical Bond,* Cornell University Press, Ithaca, 1948. 2nd edition.
46. G. Polya, *Acta Math.* **68** (1937) 145.
47. L. B. Kier, W. J. Murray, M. Randić, and L. H. Hall, *J. Pharm. Sci.* **65** (1976) 1226.
48. L. B. Kier and L. H. Hall, *ibid.* **65** (1976) 1806.
49. L. B. Kier, in: *Physical and Chemical Properties of Drugs,* S. H. Yalkowsky, A. A. Sinkula, and S. C. Valvani, Eds., Dekker, New York, 1980.
50. A. J. Stuper, W. E. Brügger, and P. C. Jurs, *Computer Assisted Studies of Chemical Structure and Biological Function,* Wiley-Interscience, New York, 1979.
51. D. R. Henry and J. H. Block, *J. Med. Chem.* **22** (1979) 465.
52. L. B. Kier and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976.
53. L. H. Hall, in: *Computational Chemical Graph Theory,* D. H. Rouvray, Ed., Nova Science Publishers, Commack, 1990, p. 201; L. B. Kier and L. H. Hall, *Reports Mol. Theory* **1** (1990) 121; L. H. Hall and L. B. Kier, in: Reviews in Computational Chemistry II, K. B. Lipkowitz and D. B. Boyd, Eds., VCH, New York, 1991, p. 367.
54. J. G. Topliss and R. J. Costello, *J. Med. Chem.* **15** (1972) 1066; R. Benigni and A. Giuliani, *Quant. Struct. — Act. Relat.* **10** (1991) 99; A. Birenzvige, L. M. Sturdivan, G. R. Famini, and P. N. Krishnan, *Comput. Chem.* **17** (1993) 33.
55. J. G. Topliss and R. P. Edwards, *J. Med. Chem.* **22** (1979) 1238.
56. G. G. Hall, *Mol. Phys.* **33** (1977) 551.
57. D. Bonchev and N. Trinajstić, *J. Chem. Phys.* **67** (1977) 4517.
58. A. T. Balaban and I. Motoc, *Math. Chem. (Mülheim/Ruhr)* **5** (1979) 197.
59. K. Altenburg, *Kolloid Z.* **178** (1961) 112.
60. E. H. Lieb and D. C. Mattis, *Phys. Rev.* **125** (1962) 164.
61. E. Ruch, *Acc. Chem. Res.* **5** (1972) 49.
62. I. Gutman and M. Randić, *Chem. Phys. Lett.* **47** (1977) 15; M. Randić, *ibid.* **55** (1978) 547.
63. R. F. Muirhead, *Proc. Edinburgh Math. Soc.* **21** (1903) 144.
64. G. H. Hardy, J. E. Littlewood, and G. Polya, *Inequalities,* University Press, Cambridge, 1934.
65. E. F. Beckenbach and R. Bellman, *Inequalities,* Springer-Verlag, Berlin, 1961.
66. M. Randić and C. L. Wilkins, *Chem. Phys. Lett.* **63** (1979) 332; J. T. Edward, *Can. J. Chem.* **58** (1980) 1897.
67. M. Randić and B. C. Gerstein, *J. Magn. Reson,* **43** (1981) 207.
68. M. Randić, *Int. J. Quantum Chem.: Quantum Biol. Symp.* **5** (1978) 245.
69. M. Randić, *J. Magn. Reson.* **39** (1980) 431.
70. T. Živković, *The Quantum Chemistry School of the U.S.S.R. Academy of Sciences*, Repino by Leningrad, December 1973.
71. I. Gutman and N. Trinajstić, *Topics Curr. Chem.* **452** (1973) 49.
72. W. C. Herndon, *Tetrahedron Lett.* (1974) 671; W. C. Herndon and M. L. Ellzey, Jr., *Tetrahedron* **31** (1975) 99; M. Randić, N. Trinajstić, and T. Živković, *J. Chem. Soc., Faraday Trans. II* **72** (1976) 244; A. J. Schwenk, W. C. Herndon, and M. L. Ellzey, Jr., *Ann. N. Y. Acad. Sci.* **319** (1979) 490; E. Heilbronner, *Math. Chem. (Mülheim/Ruhr)* **5** (1979) 105; S. S. D'Amato, B. M. Gimarc, and N. Trinajstić, *Croat. Chem. Acta* **54** (1981) 1; M. Randić, M. Baryzs, J. Nowakowski, S. Nikolić, and N. Trinajstić, *J. Mol. Struct. (Theochem)* **185** (1989) 95.
73. E. Heilbronner and T. B. Jones, *J. Amer. Chem. Soc.* **100** (1978) 6506.
74. R. Pariser and R. G. Parr, *J. Chem. Phys.* **21** (1953) 466; *ibid.* **21** (1953) 767.
75. J. A. Pople, *Trans. Faraday Soc.* **49** (1953) 1375.
76. T. Živković, N. Trinajstić, and M. Randić, *Mol. Phys.* **30** (1975) 517.

77. M. Randić and C. L. Wilkins, *J. Pys. Chem.* **83** (1979) 1525.
78. M. Randić and C. L. Wilkins, *Int. J. Quant. Chem.* **18** (1980) 1005.
79. M. Randić, G. M. Brissey, R. B. Spencer, and C. L. Wilkins, *Comput. Chem.* **3** (1979) 5.
80. F. Harary, *Graph Theory,*. Addison-Wesley, Reading, MA, 1971, 2nd printing; R. J. Wilson, *Introduction to Graph Theory,* Oliver & Boyd, Edinburgh, 1972.
81. F. S. Roberts, *Discrete Mathematical Model,* Prentice-Hall, Englewood Cliffs, 1976.
82. M. Randić, unpublished.
83. M. Randić, *J. Comput. Chem.* **1** (1980) 386.
84. M. Randić, W. L. Woodworth, and A. Graovac, *Int. J. Quantum. Chem.* **24** (1983) 435.
85. O. E. Polansky and I. Gutman, *Math. Chem. (Mülheim/Ruhr)* **5** (1979) 219.
86. C. Jordan, *J. Reine Angew. Math.* **70** (1869) 185.
87. D. Bonchev, A. T. Balaban, and O. Mekenyan, *J. Chem. Inf. Comput. Sci.* **20** (1980) 106; D. Bonchev, A. T. Balabn, and M. Randić, *J. Quantum Chem.* **19** (1981) 61.
88. D. Bonchev, O. Mekenyan, and A. T. Balaban, *J. Chem. Inbf. Comput. Sci.* **29** (1989) 91; D. Bonchev, *J. Mol. Struct. (Theochem.)* **185** (1989) 155.
89. M. Randić, *J. Math. Chem.* **4** (1990) 157; *J. Chem. Educ.* **69** (1992) 713.
90. M. Randić and N. Trinajstić, *J. Mol. Struct. (Theochem)* **284** (1993) 209; *New J. Chem.* in press.
91. N. Trinajstić, D. J. Klein, and M. Randić, *Int. J. Quantum Chem.: Quantum Chem. Symp.* **20** (1986) 699.
92. M. Randić and N. Trinajstić, *Math. Chem. (Mülheim/Ruhr)* **13** (1982) 271.
93. D. Bonchev and N. Trinajstić, *Int. J. Quantum Chem.: Quantum Chem. Symp.* **12** (1978) 293.
94. M. Randić, *J. Chem. Inf. Comput. Sci.* **32** (1992) 97.
95. W. R. Müller, K. Szymanski, J. V. Knop, Z. Mihalić, and N. Trinajstić, *J. Chem. Inf. Comput. Sci.* **33** (1993) 231.
96. M. Randić and N. Trinajstić, *J. Mol. Struct. (Theochem)* **300** (1993) 551.
97. L. Spialter, *J. Chem. Doc.* **4** (1964) 261; *Ibid.* **4** (1964) 269.
98. D. H. Rouvray, *Sci. Amer.* **254** (1986) 40.
99. W. Karcher and J. Devillers, in: *Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology,* W. Karcher and J. Devillers, Eds., Kluwer, Dordrecht, 1990, p.1.
100. G. S. Hammond, *J. Chem. Educ.* **51** (1974) 559.
101. C. A. Coulson, in a letter to MR (1973).

SAŽETAK

Graf-teorijske korelacije – izmišljotine ili činjenice?

Milan Randić, Zlatko Mihalić, Sonja Nikolić i Nenad Trinajstić

Analizirane su relacije struktura-svojstvo i različiti načini njihova prikazivanja. Predložen je skup pravila za procjenu upotrebljivosti takovih relacija. Naglašena je razlika između empirijskih, teorijskih i strukturnih korelacija, koja pomaže da se odgovori na pitanje da li su graf-teorijske korelacije izmišljotine ili činjenice. Iz provedene analize slijedi da su graf-teorijske korelacije prave korelacije između strukturnih i fizikalno-kemijskih parametara koje imaju veliku prediktivnu moć.