Journal of Mathematical Chemistry 4(1990)157–184

THE NATURE OF CHEMICAL STRUCTURE

Milan RANDIĆ

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

Abstract

We briefly discuss chemical structure as an object of mathematical characterization and list various structural invariants suitable for such a characterization. We restrict our attention to modeling compounds as graphs and examine several diverse structure–property–activity problems as an illustration of different mathematical analyses. Specifically, we consider: (1) the equivalence of an apparent quantum chemical problem as a graph theoretitcal decomposition (the analysis of diamagnetic susceptibility); (2) partial order as a tool for the illustration of regularities in isomeric variation of molecular properties (boiling points in alkanes); (3) ranking of structures as a tool in a search for biologically active substructures, illustrated on mutagenicity of nitrosamines; and (4) construction of search vectors as a tool for finding structures with a prescribed property. We end the discussion by pointing out advantages of mathematical descriptors versus physicochemical properties as descriptors. In conclusion, we focus attention on two problems facing graph theoretical approaches to chemical structure: How to incorporate differences between heteroatoms into molecular graphs, and how to incorporate spatial characteristics of chemical compounds into graph theoretical approaches. Finally, we generalize the traditional graph theoretical approaches, based on graphs and weighted graphs, to physicochemical matrices associated with molecular systems and point out the potential role that structural invariants play in discussions of molecular properties. Within this more general point of view, the quantum chemical computations produce but a fraction of the possible structural invariants that one may consider for a given system.

1. Introduction

"Chemical bond" and "chemical structure" are elementary concepts in chemistry representing the different emphasis of theoretical models of molecules. The first is concerned with the properties of electrons within the fields of the nuclei, and the pursuit of these ideas results in a particular stable (or less stable, as it may be) molecule. The second idea accepts a given connectivity and the analysis of its consequences follows. Hence, the graph theoretical approach to chemical structure is not a replacement for the quantum theoretical approach to chemical structure. On the contrary, it plays a complementary role, allowing meaningful comparisons between results and properties on different molecules or various fragmentary results on a single structure. Before we consider the similarities and differences between the two approaches, at two distinctive levels of modeling, we have to address the questions:

What is a structure?

What is the chemical structure?

2. What is a structure?

Chemistry is rich in concepts that are of practical use and interest, yet many of them have not been rigorously defined, leaving them ambiguous. It may well be that insistence on rigid formulations of some of the ideas could make them unattractive, cumbersome and less useful. Only when expectations are raised that the accumulated data can be better rationalized with a quantitative model has the time come to attempt to resolve the ambiguities associated with a qualitative concept. In table 1, we list a number of ambiguous concepts used in chemistry, starting the list with "structure". Most people have a rather good idea what chemical structure is (and what it is not), even though they would find it very difficult to qualify their notion of a structure. It may be instructive, therefore, to start with the common interpretation of such terms and then try to suitably modify this to fit the needs of chemistry. According to Webster's Dictionary [1],

Table 1

Ambiguous concepts in chemistry

Structure	Complexity
Shape	Branching
Size	Cyclicity
Similarity	Reactivity
Profile	Function

Structure is: (1) something made of parts fitted or joined together;

(2) the essential supporting portion of this;

(3) the way in which constituent parts are fitted or joined together; or arranged to give something its peculiar nature or character.

Hence, "structure" – to start with – has several distinct meanings, and while this is tolerated in linguistics, it causes ambiguities in science and is, of course, unacceptable in mathematics. Turro [2] touches on the distinction between a *graph*, as a topological object which contains information on connectivity; a *form*, a flexible topological object embedded in Euclidean space; a *figure*, a rigid geometrical object in Euclidean space; and has reserved the term *structure* for the mathematical object that models a chemical object (e.g. molecule). The preceding glossary of terms appears useful, since it reduces the possible confusion when the terms are used interchangeably.

Table 2

Mathematical objects of interest in chemistry

Ordered pair	Polytopes
Quaternions	Knots
Vectors	Lattices
Matrices	Sets
Graphs	Polynomials
Polyhedra	Sequences
Partial orders	

Observe that while the above characterization of graphs, of forms, and of figures is explicit, when it comes to "structure" the description uses auxiliary concepts: "mathematical object" and "model". This may appear as a less precise characterization, but mathematical objects are well defined (table 2 illustrates some of them) since they are constrained to (suitable) mathematical operations, which are always rigorously defined. The concept of a "model" is also generally well understood, and accepted as an alternative form of an object in which specific details of interest are emphasized so that their study may offer some insight into the nature of the object itself.

Hence, in summary, we can say that structure is an object built from components. The components can be selected such as to give particular emphasis (the second meaning in Webster's definition) and to offer particular insight, and they themselves represent mathematical objects. Combining components in a structure can be associated with the term "pattern" ("the way in which . . .", the third meaning of the Webster definition). If we adopt this pragmatic approach to a structure, and then to chemical structure in particular, observe that we still have a great lattitude in selecting *components* which represent a structure; and even after components have been selected, we have considerable lattitude in selecting mathematical descriptors for such components. The chemical structure can therefore be viewed as constructed from atoms and bonds (as most common models assume), it can be viewed as composed of fused rings, it can be viewed as composed of a collection of substructures (which can be, for example, Kekulé valence structures in the case of conjugated benzenoids), the building blocks can be fundamental bases (as in proteins), etc. On the other hand, mathematical objects of interest may be matrices, polynomials, sequences, sets, partial orders, lattices, etc. In table 3 we list numerous components which may be viewed as building blocks of a chemical structure while, similarly, in table 4 we give a short list of mathematical objects of potential interest to represent structure and their components. Observe that components that contribute to the characterization of a structure are not necessarily molecular fragments (such as atoms, bonds, group of atoms, rings, etc.), but include more abstract elements such as hybrid orbitals, molecular orbitals, Kekulé valence structures, conjugated circuits [3], etc.

Table 3

Components of a molecular structure

Atoms	Kekulé valence structures
Bonds	Conjugated circuits
Paths (of different length)	Bond orbitals
Fragments	Non-adjacent numbers
Rings	Bond types
Atomic orbitals	

Table 4

Objects, subject to mathematical manipulation, of interest for the characterization of molecular structure

Sets	Sequences
Kekulé structures	Path numbers
Atomic ID numbers	Characteristic polynomial
Ring indices	Acyclic (matching) polynomial
Ulam's subgraphs	Self-returning walks
Bond orders	Random walks
Atomic charges	Conjugated circuits
	Weighted paths

Important possibilities that the preceding pragmatic approach lead to are:

(a) Lists of components together with *instructions* of how these are combined to give the structure. We refer to such as a *molecular code*, based on the particular components.

(b) Sets of components, i.e. collections of components for each structure. We refer to these as a *projection* of a structure. Observe that we lose information with a "composition rule", and it may not be clear how to reconstruct the structure from its projection. It is an open problem whether a structure can be recovered from a sufficiently large number of such projections.

(c) Naturally *ordered* sets of components, to be referred to as *descriptions*. Individual elements of natural sequences are referred to as molecular descriptors.

(d) Finally, structure can be represented by a single number which captures much of the structural features. We will refer to these as *topological* indices. If such an index maintains a high discriminatory power, it may be referred to as a molecular ID number [4].

The list of components that provide a basis for the characterization of structures, exemplified in table 3, is by no means complete, and equally there is no restriction on construction of yet unconsidered invariants. The ultimate criterion in judging pragmatic descriptions is the utility.

We have separated the mathematical objects of table 4 into two groups, one corresponding to "molecular projections" and the other to "molecular descriptions". Explanations of what is meant by "natural" order and what is meant by "capturing" structural features are needed in order to clarify the above classification.

"Natural" ordering implies some correspondence with the "size" of the components. For example, path numbers of different lengths allow one to order the count of such paths naturally. Similarly, conjugated circuits occur with different sizes and can be simply ordered. However, there is no such apparent natural way of ordering Kekulé structures or Ulam's subgraphs (i.e. subgraphs obtained by erasing one of the vertices, each time a different one). Orderings based on numerical (computational or empirical) procedures are generally excluded because of their dependence on a particular parametrization adopted, and such ordering may change with reparametrizations. Observe that a sequence implies a *complete* order, not a *partial* order which typically emerges in a comparison of structures. Partial order is typified with "ties", i.e. two or more structures being noncomparable, and thus allowing different complete orders. The difficulty with an "excessive" basis, such as that of all subgraphs, or even of all Ulam's subgraphs, is that the number of components grows quickly and their use becomes impractical when considering the characterization of larger structures.

By "capturing" essential structural features, we mean that such severe projections (of a structure to a single number) should satisfy the basic requirement that apparently similar structures should be associated with numerically similar (i.e. not widely different) numbers. This is the case with a number of topological indices, such as Hosoya's Z (based on the count of nonadjacent bonds [5]), the connectivity index X (based on discrimination of (m, n) bond types [6]), the molecular ID number [4], and many other indices.

3. Diamagnetic susceptibilities

Hameka considered a quantum chemical model for computing the diamagnetic susceptibilities in alkanes [7] and other organic compounds [8]. Using an MO approach in the case of alkanes, he introduced the following plausible assumptions:

(1) CC and CH bonds are localized.

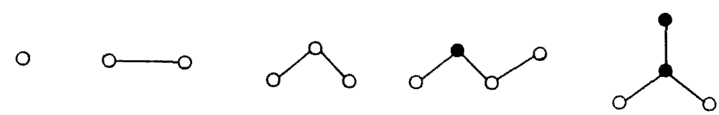
(2) All CC and CH bonds are the same, for all alkanes.

A summary of Hameka's analysis is given in table 5, which gives the diamagnetic susceptibilities as additive in terms of the following three parameters:

Table 5

Decomposition of the contributions to magnetic susceptibilities in alkanes according to quantum chemical MO modeling (capital labels) and graph theoretical modeling (small labels)

Molecule	Hameka	Graph theory
methane	$A + C$	a
ethane	$A - B$	$2a + b$
propane	$A + 2B$	$3a + 2b + c$
butane	$A + 3B$	$4a + 3b + 2c + d$
isopropane	$A + 3B + C$	$4a + 3b + 3c + e$
pentane	$A + 4B$	$5a + 4b + 3c + 2d$
2-methylbutane	$A + 4B + C$	$5a + 4b + 4c + 2d + e$
neopentane	$A + 4B + 3C$	$5a + 4b + 6c + 4e$
hexane	$A + 5B$	$6a + 5b + 4c + 3d$
2-methlypentane	$A + 5B + C$	$6a + 5b + 5c + 3d + e$
3-methlypentane	$A + 5B + C$	$6a + 5b + 5c + 4d + e$
2,2-dimethylbutane	$A + 5B + 3C$	$6a + 5b + 7c + 3d + 4e$
2,3-dimethlybutane	$A + 5B + 2C$	$6a + 5b + 6c + 4d + 2e$



a *b* *c* *d* *e*

$A = a - c$ $\quad a = A + C$

$B = a + b + c$ $\quad b = -A + B - 2C$

$C = c$ $\quad c = C$

$$A = X(\mathrm{C}) + 4X(\mathrm{CH}) + X(\mathrm{CC,CC}) - 2X(\mathrm{CC,CH}) - 5X(\mathrm{CH,CH});$$

$$B = X(\mathrm{C}) + X(\mathrm{CC}) + 2X(\mathrm{CH}) - X(\mathrm{CC,CC}) - 4X(\mathrm{CC,CH}) - X(\mathrm{CH,CH});$$

$$C = -X(\mathrm{CC,CC}) + 2X(\mathrm{CC,CH}) - X(\mathrm{CH,CH}).$$

If bonds are localized, identical, and constant within a set of compounds, then one could say the connectivity is of essence, not the fine details of the individual bonds (which the model ignores). This then means that the problem is a "disguised" graph theoretical problem and can be cast in an equivalent graph theoretical formulation. By considering the hydrogen suppressed graphs of methane, ethane, and propane (i.e. vertices (a), edges (b), and paths of length two (c), corresponding to the propane carbon skeleton) and by decomposing other alkanes in terms of a, b, c, we immediately obtain equivalent results (shown in table 5) with simple equivalence relations [9]:

$$A = a - c; \quad B = a + b + c; \quad \text{and} \quad C = c.$$

Graph theoretical analysis has not "discovered" something new: it was Hameka and his quantum chemical modeling which led to the recognition of the atom and bond additivities for the diamagnetic susceptibilities. The graph theoretical approach is, however, more "transparent" and, perhaps, more clearly points to the components which are essential in this particular molecular additivity.

4. Ordering of isomers

Consider fig. 1, listing the eighteen isomers of octane and their boiling points. Is there some regularity between the structural forms and the relative magnitudes for the boiling points?

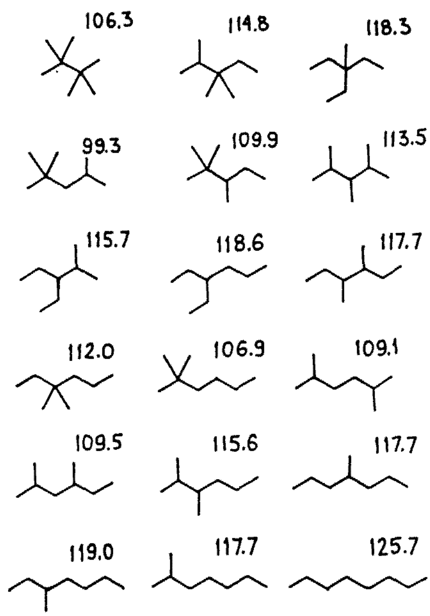


Fig. 1. The octane isomers and their boiling points. Is there any regularity in this figure?

Perhaps one sees some trends, but admittedly it is not apparent that there is a regularity; even less what it is one should look for! This problem was considered from the graph theoretical point of view, which can be reformulated as: In what structural components do the eighteen isomers differ? Only if we identify such discriminatory descriptors can we compare individual structures and hope to see if a particular descriptor can account for the observed isomeric variations. Isomers have the same number of carbon atoms (and the same number of hydrogens) and the same number of CC bonds; hence, we have to go *beyond* atoms and bonds in a search for useful descriptors. There are no obvious directions in which to go, i.e. in which way to augment the model. For example, the next important discriminator could be the number of close atomic "contacts", which will depend on the stereochemistry of the compounds, the next structural element could equally be the number of possible rotamers, which is different for different isomers, or possibly the role of the next-nearest-neighbor interactions, etc. If, however, we decompose the octanes in a fashion similar to that when examining their diamagnetic susceptibilities, and include larger path components, we arrive immediately at the rather simple mathematical descriptors listed in table 6 [10].

Table 6

Isomers of octane and their path numbers

	Molecule	Number of paths of length i						
		p_1	p_2	p_3	p_4	p_5	p_6	p_7
A	2,2,3,3-tetramethylbutane	7	12	9				
B	2,2,4-trimethylpentane	7	10	5	6			
C	2,2,3-trimethylpentane	7	10	8	3			
D	2,3,3-trimethylpentane	7	10	9	2			
E	2,3,4-trimethylpentane	7	9	8	4			
F	2,2-dimethylhexane	7	9	5	4	3		
G	3,3-dimethylhexane	7	9	7	4	1		
H	2,5-dimethylhexane	7	8	5	4	4		
I	2,4-dimethylhexane	7	8	6	5	2		
J	2,3-dimethylhexane	7	8	7	4	5	2	
K	3-methyl-3-ethylpentane	7	9	9	3			
L	2-methyl-3-ethylpentane	7	8	8	5			
M	3,4-dimethylhexane	7	8	8	4	1		
N	2-methylheptane	7	7	5	4	3	2	
O	3-methylheptane	7	7	6	4	3	1	
P	4-methylheptane	7	7	6	5	2	1	
Q	3-ethylhexane	7	7	7	5	2		
R	*n*-octane	7	6	5	4	3	2	1

We see that, even though there are no guarantees that the particular descriptors will produce a unique characterization, there is an apparent high discrimination between the isomers when path numbers are used as descriptors. If we restrict our attention to the leading path numbers in which octane isomers differ, p_2 and p_3, we still maintain a high discrimination between the individual octanes. By viewing (p_2, p_3) as coordinates, we can arrange all octanes as shown in fig. 2 [12]. We can replace coordinate sites with "blocks" and obtain a table reminiscent of the Periodic Table of Elements, with the individual sites in the table corresponding to isomers. By inserting experimental

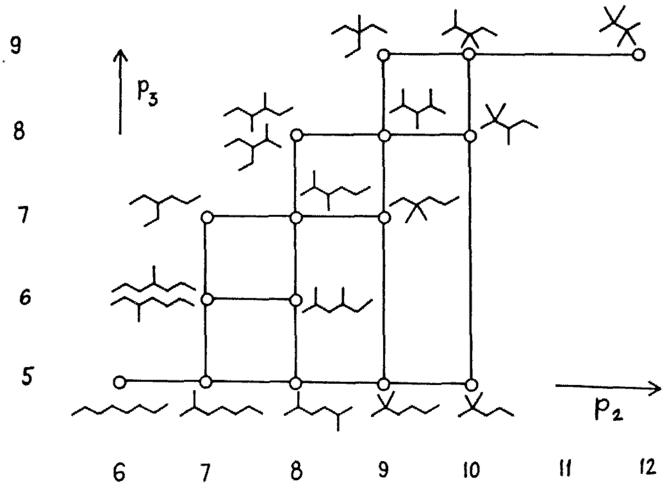


Fig. 2. Octane isomers arranged on a p_2, p_3 coordinate grid.

n-8 125.7				
2M 117.7	3M 4M 119.0 117.7	3E 118.6		
2,5 MM 109.1	2,4 MM 109.5	2,3 MM 115.6	2M3E 3,4 MM 115.7 117.7	
2,2 MM 106.9		3,3 MM 112.0	2,3,4 MMM 113.5	3M3E 118.3
2,2,4 MMM 99.3			2,2,3 MMM 109.9	2,3,3 MMM 114.8
				2,2,3,3- MMMM 106.3

Fig. 3. The isomers of fig. 2 depicted as a Table of Isomers (p_2 as columns and p_3 as rows) with the boiling points inserted to illustrate a regular variation in the BP with p_2 and p_3.

properties, one can discern regularities, as is illustrated by the boiling points in fig. 3. One immediately sees a simple regularity among the boiling points of octane isomers: the relative magnitudes decrease with an increase in p_2, and somewhat increase with an increase in p_3. Not only have we in this way found, using the mathemtical properties of graphs, the regularity sought, but as further studies show [12–15], such a Table of Isomers, as we can rightly refer to fig. 3, will show regularities in other physicochemical properties of alkanes. This includes properties that do not correlate among themselves, such as indices of refraction and liquid densities of alkanes, which do not correlate with the boiling points, heats of formation, molar volumes, etc. Moreover, the same Table of Isomers can be used to suggest the construction of novel "molecular" properties derived by combining selected atomic properties, as has been illustrated for 13-C chemical shifts in alkanes [13,15].

5. Search for the critical substructure

In fig. 4, we depict thirteen nitrosamines together with their relative mutagenic activities (as reported in the literature [16]). The structures are labeled alphabetically relative to their decreasing mutagenicity potential. The question here we would like to

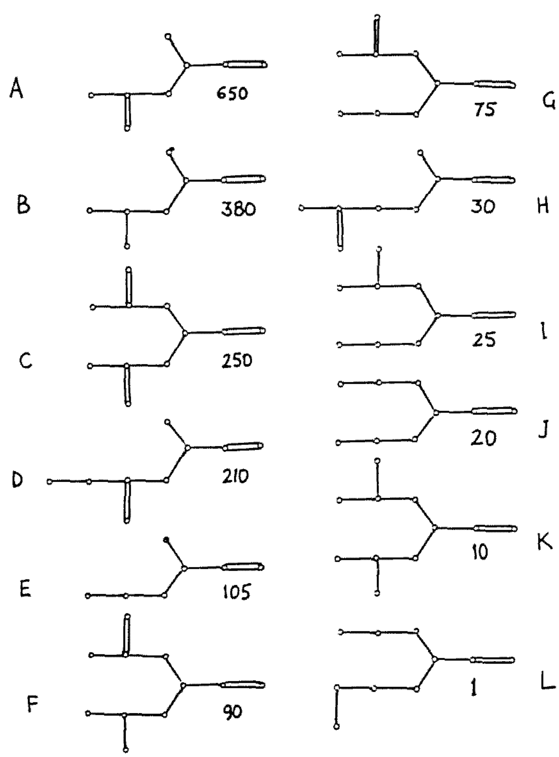


Fig. 4. Nitrosamines with their relative mutagenicity potentials. Is there some regularity here?

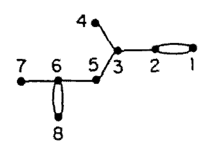
ask is: "Why do these, apparently similar compounds, show such large differences in their mutagenicity?" Or, in other words: "What is the underlying regularity in fig. 4?". Which structural factor is responsible for the observed relative differences in the mutagenicities of nitrosamines which cover the range of more than three orders of magnitude?

This problem is different to that of the ordering of octanes in fig. 1 in that the molecules here are of different sizes and differ in the presence or absence of some substituents. Accumulated experience in medicinal chemistry and pharmaceutical studies suggests that some active *substructure* may be responsible for the bioactivity, and our task is then to identify such a substructure. A useful tool to examine molecular fragments is that of atomic ID numbers, which are sums of the weighted paths emanating from an individual atom [4]. Atomic ID numbers offer a basis for the construction of various "fragment projections". They represent a collection of invariants from which one can choose those belonging to specified fragments. In this way, different fragments can be studied and, hopefully, the pertinent one identified.

We start with weighted paths, and in table 7 illustrate the computer output of the ALLPATH program on methyl-2-oxypropylnitrosamine (structure A), the most potent mutagen among those of fig. 4. Because in all compounds considered the heteroatoms

Table 7

Weighted path numbers for a nitrosamine of fig. 4 (compound A)



Atom	P_0	P_1	P_2	P_3	P_4	P_5	Atomic ID
1	1	0.8164	0.2721	0.2682	0.0392	0.0474	2.4436
2	1	1.1498	0.3285	0.0481	0.0580		2.5845
3	1	1.3189	0.4165	0.1742			2.9096
4	1	0.5773	0.4281	0.2404	0.1005		2.3465
5	1	0.7618	0.7985	0.1111			2.6714
6	1	1.5606	0.1443	0.1314	0.03928		2.8757
7	1	0.5000	0.5303	0.0721	0.0657	0.0196	2.1878
8	1	0.7071	0.6035	0.1020	0.0929	0.0277	2.5334
Molecule:	8	3.6960	1.7610	0.5739	0.1979	0.0474	14.2764
	Atom count	Molecular connectivity		Higher connectivities			Molecular ID

are in "fixed" positions, we represented molecules by graphs without differentiating heteroatoms. The weighting procedure is based on assigning to a bond type (m, n), where m and n indicate numbers of nearest-neighbors (i.e. the valencies of the two vertices in the edge), the weight 1/sqrt (m, n), just as used in the design of the connectivity index [6]. Atomic ID numbers are shown in the last column, and these are the numbers used to represent the individual compounds considered. Because in this case it is not difficult to recognize the "corresponding" atoms in different compounds, ordered sets on n-tuples could be viewed as vectors in n-dimensional Euclidean space. Thus, if we select a 7-atom fragment (associated with labels 1–7 in table 8), we obtain for the first compound:

A: (2.44, 2.58, 2.90, 2.34, 2.67, 2.87, 2.18).

The above are truncated entries of the last column in table 7. Similarly, one finds that the second compound is characterized by the projection [18]:

B: (2.45, 2.59, 2.95, 2.36, 2.77, 2.92, 2.36).

Table 8

Atomic ID numbers for the seven atoms forming the fragment to be investigated[a] (numbering of atoms as in table 7)

	Atom Molecule	1	2	3	4	5	6	7
A	MOP	2.443	2.584	2.909	2.346	2.671	2.875	2.187
B	MHP	2.454	2.598	2.950	2.369	2.770	2.926	2.356
	DMN	2.402	2.534	2.760	2.260	2.260	–	–
C	BOP	2.484	2.634	3.054	2.732	2.732	2.897	2.198
D	2-MOB	2.447	2.584	2.924	2.355	2.708	2.979	2.547
E	MP	2.451	2.594	2.939	2.363	2.744	2.652	2.375
F	HPOP	2.495	2.647	3.099	2.749	2.831	2.903	–2.201
							2.951	–2.370
G	POP	2.492	2.644	3.089	2.744	2.805	2.683	2.397
								2.200
H	3-MOB	2.455	2.594	2.594	2.372	2.781	2.725	2.894
I	2-HPP	2.503	2.657	3.129	2.844	2.822	2.691	2.403
								2.373
J	DP	2.500	2.654	3.118	2.817	2.817	2.684	2.401
K	BHP	2.506	2.661	3.140	2.848	2.848	2.958	2.374
L	3-HPP	2.508	2.664	3.148	2.830	2.891	2.835	2.698
								2.405

[a]Only data for the seven common nonhydrogen atoms are represented. In some cases, there are two alternative choices for the seven atoms and both alternatives are shown.

Similar results follow for other nitrosamines of fig. 4, which are listed in table 8. It is not difficult now to evaluate the "distance" between any pair of structures. It suffices, however, to consider the structures A and B as the standards and evaluate the distances between the other structures and the standards. The standards A and B are the compounds with the greatest bioactivity. The derived distances give a measure of similarity between the compounds, based on the selected 7-atom molecular fragment. With respect to A, we obtain [17]:

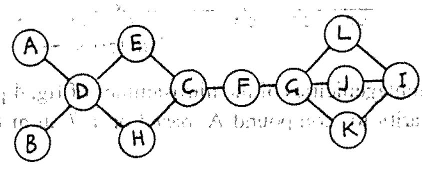
B	D	C	F	G	K	E	J	I	L	H
0.20	0.37	0.42	0.48	0.50	0.54	0.54	0.61	0.63	0.63	0.80,

and with respect to B, we obtain a different ordering:

A	D	E	C	F	G	L	J	K	I	H
0.20	0.21	0.27	0.41	0.41	0.47	0.53	0.54	0.55	0.56	0.57.

From the above two orderings, one can extract a *partial order*, i.e. a set of all fragmentary orders that are present in *both* the above sequences. For example: the sequence D, G, J, I, H is one such subsequence because, as one can see upon inspection, in *both* sequences this partial order is contained. A simple way to extract all such embedded partial orders is illustrated at the top of fig. 5, where the two sequences, based on the leading structures A and B as standards, were written one above the other. Subsequently, the same labels in the two sequences are connected. Each crossing of a line indicates a pair of structures which have an inverted order and hence cannot be "compared", i.e. do not dominate each other. Hence, a label (structure) which does not cross lines of other labels (compounds) dominates all such compounds if it is closer to the standard. In this way, we derive the pictorial representation of the particular partial order, shown in the middle of fig. 5.

The test of the assumption that the particular 7-atom fragment is responsible for the relative mutagenicity consists of replacing structures in the derived partial order with the numerical values for the mutagenicity and examining if such a replacement involves serious contradictions. A contradiction is reflected by a reversed order of the relative magnitudes of mutagenicity, while acceptable results should be accompanied with a regular decrease of the numerical values for the property (mutagenicity) as we move from the standards at the left to less and less similar compounds at the right-hand side of the diagram. From fig. 5, we can conclude that there are no serious contradictions in our result, even though the particular partial order has a few minor discrepancies. Since we labeled the compounds alphabetically, unacceptable results would appear as a partial order, which seriously violates an alphabetical arrangement associated with the hierarchical graph of the partial order. For example, when we consider the fragment to consist of six atoms only (eliminating atom 7), we obtain the following diagram for the partial order:



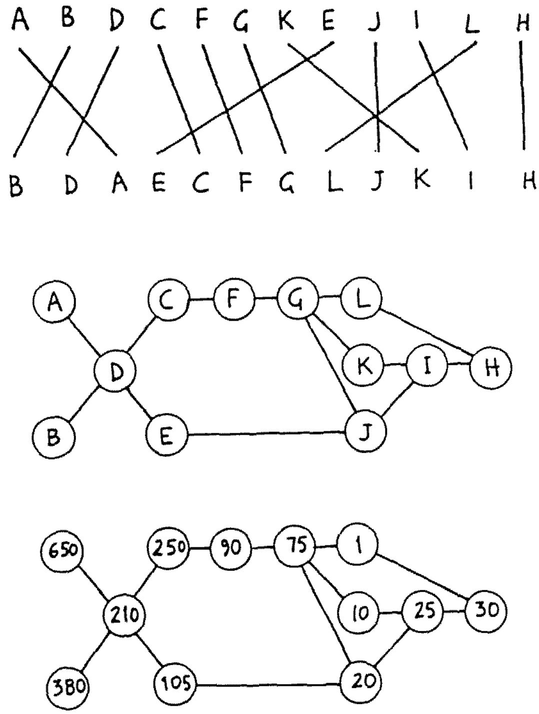


Fig. 5. Partial order derived for nitrosamines of fig. 4 (ordered in decreasing similarity to the standards A and B), when a 7-atom fragment is taken as a basis for comparison and similarity is based on (weighted) atomic ID numbers.

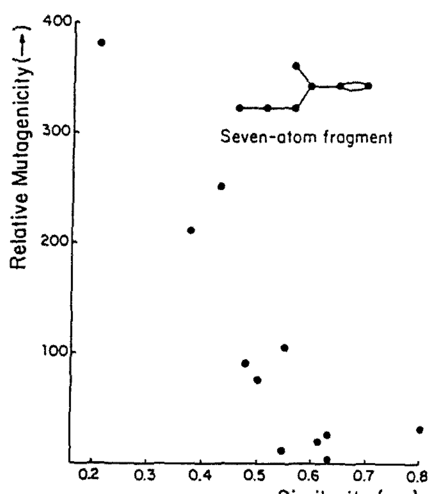


Fig. 6. Relative mutagenicities of the nitrosamines of fig. 4 plotted against decreasing similarity to compound A, based on a 7-atom fragment.

which shows several apparent alphabetical "conflicts". To see that the result of fig. 5 is indeed acceptable (despite its own minor discrepancies) or, in other words, to recognize the discrepancies in the diagram (fig. 5) as minor, we depict in fig. 6 the correlation between the reported mutagenicities and the similarity of the nitrosamines with respect to A (based on the numerical values in the sequence previously shown). Theoretically, one expects the regression to be given by some descending function. Indeed, as we see from fig. 5, there is a high correlation between the bioactivity and the degree of similarity (based on 7-atom fragment and atomic ID as mathematical characterization).

6. Design of a structure with desired property

The capability of graph theoretical approaches to discern the essential parts of molecules to be associated with a particular property is of great interest. The example illustrated a *direct* quantitative scheme for such explorations. Previous approaches to the detection of important fragments, such as implied in the "morphine rule" [18], were based on chemical intuition and experience, or alternatively on statistical inference [19]. In this section, we will briefly illustrate how to use the outlined methodology in a search for a compound of desired property and how to verify that a particular compound belonging to a certain class is the best. This will be achieved without screening all compounds of a given class, which would be difficult even for families with a relatively small number of compounds.

Consider the nine benzomorphans of fig. 7, which have been alphabetically labeled in parallel to their relative analgesic activities. Again, we should search for a molecular fragment which is the basis for comparisons of the molecular properties and

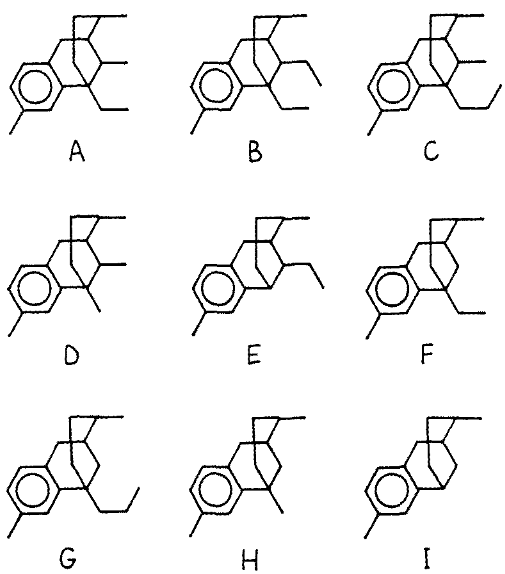


Fig. 7. Benzomorphans arranged alphabetically with decreasing analgesic potency.

which, if found, would produce the illustrated, alphabetical ordering of the compounds. We test the same fragment in different molecules against the same fragment in the standard, benzomorphan A, the most potent analgesic. Preliminary examinations of the benzomorphans [20] included the largest common substructure with $n = 15$ atoms as the fragment in the nine benzomorphans considered, the "morphine rule" fragment of $n = 11$ atoms, and a smaller fragment with $n = 8$ atoms, illustrated in fig. 8. All the fragments considered could account qualitatively for the observed ordering of the benzomorphans quite well.

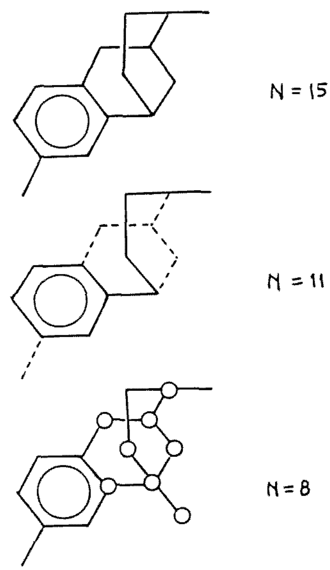


Fig. 8. Various fragments of interest in a search for the "best" benzomorphan compound.

We will now consider the reverse problem: We will pretend that we do not know that A is the best compound among the nine benzomorphans and will see if, by some *systematic*, objective procedure, we can point to A as the best compound. We selected the smallest fragment of fig. 8 ($n = 8$) to illustrate the search for an optimal structure. Suppose that at random we synthesized compound E, in the middle of the group, which upon testing shows a fair bioactivity. Then, in a similar fashion, we obtain another compound in the class, say D, which shows a better performance. Were the second compound found, compound F rather than D, which signals a weaker bioactivity, we could discard it from further consideration because it points in a "wrong" direction from our search. However, even such negative results could be used, as a characterization of the opposite (undesirable) direction. However, in order to simplify the outline of the search procedure, let us consider how to use the novel compound D, which shows a better characteristic, with the information on the previously obtained compound E. The two structures E and D can be "averaged", producing a hypothetical structure

(D + E)/2 for which we can write the molecular "projection", the set of atomic ID numbers that such a hypothetical structure would possess. The search continues, and upon hitting the compounds C and B we should in the same way construct hypothetical average structures (C + D + E)/3 and (B + C + D + E)/4, respectively. The effect of the averaging is an enhancement of the structural characteristic that is critical for the particular property and attenuation of those structural features in which the compounds differ, and which presumably therefore cannot be responsible for the particular property. In table 9, we show the atomic ID numbers for the common eight atoms of the hypothetical structures, together with the extrapolated atomic ID values obtained

Table 9

Atomic ID numbers for hypothetical structures obtained by averaging structures with higher activity with the extrapolated "best" atomic ID descriptors approaching the "missing" structure A

Atom	E	(D + E)/2	(C + D + E)/3	(B + C + D + E)/4	Extrapolation	A
1	0.365	0.367	0.363	0.361	0.359	0.360
2	0.383	0.384	0.380	0.379	0.375	0.376
3	0.378	0.374	0.369	0.370	0.363	0.364
4	0.405	0.406	0.406	0.406	0.407	0.407
5	0.286	0.287	0.298	0.295	0.305	0.308
6	0.338	0.339	0.337	0.336	0.336	0.336
7	0.342	0.343	0.340	0.338	0.336	0.337
8	0.396	0.398	0.395	0.393	0.392	0.393
Euclidean distance from A	0.029	0.028	0.013	0.015	0.004	–

by a least-squares linear fit of the four points, corresponding to E, (D + E)/2, (C + D + E)/3, and (B + C + D + E)/4, taken as guiding us toward the structure of desired optimal activity. The last column in table 9 reproduces the atomic ID for the eight atoms of the leading structure A, the "best" compound in the class. We immediately see that an extrapolation leads to a hypothetical structure A′ with the smallest difference with structure A (measured by viewing structures as vectors in an 8-dimensional Euclidean vector space). If we systematically examine all derivatives of benzomorphan of interest by exploring all available substitution sites and compare the corresponding vectors based on the atomic ID values, we could in principle exhaust the pool of reasonable structures and with certainty deduce which is the best structure, i.e. which structure is at the smallest distance from A′. The idea of "search" vectors has been advanced by Venkataraghavan and collaborators [21] in combination with their own topological descriptors. Their search vectors were defined, not as we outlined here by an "iterative" (and interactive) procedure, but as vectors defined between the centroids of active and the centroids of inactive compounds. There is no doubt that the search vectors, which indicate figuratively in which direction to "navigate" through the "sea" of structures makes this search more efficient. The concept is analogous to the use of a steepest gradient in "greedy" algorithms and other numerical computational problems. One can further refine such approaches by restricting the number of compounds included in the "averaging". Thus, when a new compound which qualifies for inclusion is found (i.e. a compound exceeding in property the compounds already considered), the least active compound in the averaging set is left out. This is likely to increase the signal-to-noise ratio in the search, i.e. decrease the role of irrelevant parts of the structures considered.

7. Anatomy of QSAR

Quantitative structure–activity relationships (QSAR) may be categorized as structure-cryptic empirical schemes (e.g. Hansch-type analysis [22]), structure-implicit methods (e.g. quantum chemical computations), and structure-explicit (graph theoretical) approaches [23]. Briefly, structure-cryptic methods use (often large numbers of) molecular properties as descriptors and, in fact, if successful, represent a property–property relationship rather than a structure–property relationship. Structure-implicit approaches treat the molecule as a whole, not as composed of components which, when considered, are identified in an intuitive or empirical way, and are not an integral part of the computational method. Finally, structure-explicit methods are those of graph theory [24] in which well-defined structural (mostly graph) invariants form a basis for comparisons among molecules. Although most investigators employ one of the diverse methodologies, the structure–activity phenomena are so complex that all the currently available methodologies may not suffice. Hence, one should try to combine "different" points of view and different methodologies when possible. It would facilitate further development if at the same time researchers using different schemes point out applications and individual structures which fail to agree with a particular regression or do not fit a particular description. The so-called "outliers" of a correlation, which one tends to exclude, ought to be closely examined and the structural basis for their apparent ill-behavior understood. Thence, they may lead to valuable information which may eventually point to an important improvement of the underlying model.

Most traditional QSAR reports use large numbers of experimental quantities, such as $\log P$ (partition coefficient for octanol/water as an indicator of transport characteristics), hydrophobic constants πi, $\Delta pK(a)$ values, molar volumes, molar refractivities, etc. These are then also occasionally combined with some quantum chemically computed parameters, such as HOMO and LUMO parameters that are suggestive of electron "mobility", etc. As a rule, such traditional QSAR studies avoid the use of graph theoretical descriptors, such as various connectivity indices. Because of this situation, it is in order to illustrate an example of the differences and similarities between the "traditional" approaches to QSAR and the "novel" approaches to QSAR based on graph theoretical descriptors.

In fig. 9, we show eighteen molecular skeletons of the variable fragments of 2-(arylinino)imidazolidines:

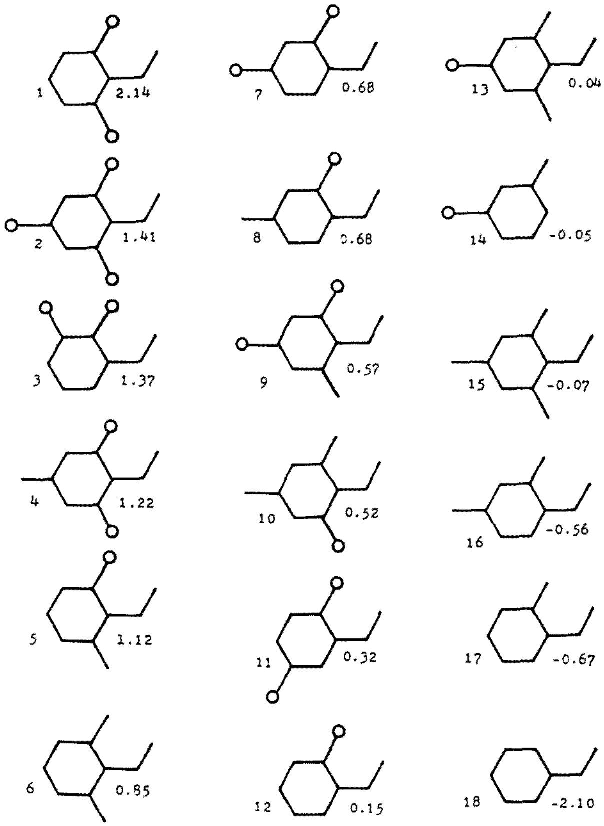
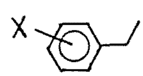


Fig. 9. Variable fragment in imidazolidines examined with experimentally reported activities (as log $1/ED$). Chlorine atoms are represented by circles.



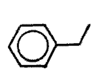
in which X is either methyl or chlorine. These compounds have known hypotensive activities and the data used are as reported by Timmermans and van Zwieten [25], whose results we re-investigated. These authors examined almost a dozen different property-based molecular descriptors, including $\log P$, $\Delta pK(a)$, the lowest electronic excitation energies as derived from the difference in HOMO and LUMO, the hydrophobic constant πi, parachor (a parameter governed by volume and surface tension of a molecule), steric parameters (as proposed by Taft), and molar refractivities. They considered a total of twenty-seven compounds, of which eighteen involve chlorine as the only heteroatom. The additional nine compounds included, besides chlorine, also fluorine, bromine, oxygen and nitrogen. We excluded these nine, being too few compounds to allow one to estimate the empirical parameters associated with these heteroatoms. Because we reduced the initial sample size, we had to re-evaluate the original correlation equations of Timmermans and van Zwieten. We arrive at the following replacements [26]. The best reported correlation based on a 5-descriptor (parachor and parachor-squared, $\Delta pK(a)$, HOMO and another MO-derived parameter) now gives $r = 0.964$ for the correlation coefficient and $s = 0.301$ for the standard deviation. These r and s values, as expected, have been improved slightly when compared to those for the set of $n = 27$ structures. The best single parameter correlation (based on $\log P$) now has $r = 0.529$ and $s = 0.864$, while stepwise inclusion of the descriptors in the best 5-parameter correlation gradually increases the correlation coefficient from $r = 0.675$ for single variable ($\Delta pK(a)$) to $r = 0.731$ for two-parameter regression (quadratic in parachor) and then further to $r = 0.902$ when the three parameters mentioned above are combined.

Let us now consider a graph theoretical approach to the *same* set of compounds. The first task is that of deciding how to discriminate between carbon atoms and chlorine. The compounds of fig. 9 have a heteroatom "floating" around, and it is no longer possible to ignore the differences between carbon atoms and heteroatoms. For example, we may compare the reported $\log 1/ED$ for the following:

2-Cl, 4-Me	2,4-Cl(2)	2-Me, 4-Cl	2,4-Me(2)	
53	61	275	810	(μg/kg),

all of which, if one does not discriminate the heteroatom, correspond to the same molecular graph. Observe also that from the $\log(1/ED)$ values shown for the four compounds, no simple "additivity" is apparent. Moreover, even the trend associated with heteroatoms is not apparent, because the di-chloro substituent is "flanked" by mono-chloro substituents. Kier and Hall's valency connectivity indices [27], because of their inherent bond additivity, are not suitable for describing the above "irregular" behavior. We therefore decided to consider the diagonal elements of the adjacency matrices as a route to discrimination among heteroatoms. Such "modifications" of the diagonal entries in the adjacency matrix correspond to what one normally does in empirical MO methods. Even in the early applications of graph theory to chemistry, as discussed at length already in 1940 by Balandin [28], atomic labels were considered as the entries on the diagonals for various matrices associated with molecular properties. The same idea can also be found later in the work of Spialter on chemical documentation [29].

Our preliminary studies [26] suggest a negative value of –0.20 as "sufficient" (but in no way an optimal value) to discriminate between chlorine and carbons (the latter, of course, have diagonal elements zero). Therefore, using for chlorine –0.20 as the diagonal matrix entry and confining the summation in deriving the connectivity indices to the eight common atoms:



present in all of the eighteen fragments, we obtain for the weighted path counts, which correspond to the leading connectivity indices, the values shown in table 10. These connectivity indices, or to be more precise, 8-atom fragment weighted paths of length 1, 2 and 3, are subsequently used in multiple regression analysis.

Table 10

The connectivity index X and higher path numbers obtained using empirical parameters (as the diagonal element for chlorine) to discriminate heteroatoms

	Compound	$1-X$	$2-X$	$3-X$
1	2,6-Cl_2	4.278	2.015	0.978
2	2,4,6-Cl_3	4.418	2.120	0.969
3	2,3-Cl_2	4.278	2.015	0.963
4	2,6-Cl_2-4-Me	4.384	2.092	0.957
5	2-Cl-6-Me	4.244	1.989	0.964
6	2,6-Me_2	4.210	1.964	0.949
7	2,4-Cl_2	4.262	2.036	0.982
8	2-Cl-4-Me	4.228	2.008	0.970
9	2,4-Cl_2-6-Me	4.384	2.095	0.955
10	2,4-Me_2-6-Cl	4.350	2.067	0.944
11	2,5-Cl_2	4.262	2.036	0.982
12	2-Cl	4.122	1.939	0.982
13	2,6-Me_2-4-Cl	4.350	2.069	0.942
14	2-Me-4-Cl	4.228	2.011	0.968
15	2,4,6-Me_3	4.316	2.042	0.931
16	2,4-Me_2	4.194	1.983	0.956
17	2-Me	4.088	1.914	0.966
18	unsubstituted	3.966	1.869	0.979

Use of a single graph theoretical parameter, the $1-X$ connectivity index derived from the modified adjacency matrix, using the standard ALLPATH program supplemented with weighting factors that have already been mentioned, gives the regression:

$$\log(1/ED) = 5.781X - 24.1643,$$

with $r = 0.690$ and standard deviation $s = 0.712$. This is visibly better than the *single* parameter correlation using molecular properties ($\log P$) or the best single parameter ($\Delta pK(a)$). Continuing, by adding $2-X$ to $1-X$, we obtain the following two-parameter regression:

$$\log(1/ED) = 20.830X - 26.636(2-X) - 34.516,$$

with $r = 0.781$ and $s = 0.635$. The improvement is considerable, though not dramatic. In part, this is because the two graph theoretical parameters $1-X$ and $2-X$ themselves have a correlation ($r = 0.608$), implying some "duplication". Nevertheless, the above two-term correlation is very comparable to the best two-parameter property-based correlation, the quadratic correlation in $\log P$ with $r = 0.786$ and $s = 0.629$. If, however, we continue and include $3-X$, we obtain the correlation equation:

$$\log(1/ED) = 36.265X - 49.192(2-X) + 46.829(3-X) - 99.830,$$

with an impressive improvement in the correlation coefficient ($r = 0.977$) and impressive reduction in the standard deviation ($s = 0.224$).

As one immediately sees, the above three-parameter graph theoretical correlation is *better* then the best *five*-parameter correlation based on physical properties, combined with quantum chemical quantities as descriptors. We should also add that the *five*-parameter best correlation of Timmermans and van Zwieten was derived after screening numerous alternatives and using some dozen property or quantum-chemically computed descriptors. In contrast, we have not used statistical analysis to select the best combination of the connectivity indices, we simply took the leading three indices. Moreover, we have not even attempted to optimize the "diagonal" entry for chlorine, taken as -0.20, since we wanted to illustrate the *flexibilities* of graph theoretical schemes rather than focusing on a search for the best correlation. If, for example, one decreases the "diagonal" parameter to -0.40 with a single connectivity index X, one increases the correlation coefficient from $r = 0.690$ to $r = 0.750$. The so improved correlation, if compared with the single-parameter traditional QSAR derived from data by Timmermans and van Zwieten, *doubles* the variance of the correlation based on $\log P$; and again, -0.40 is not the optimal value for the diagonal entry of chlorine either!

In summary, graph theoretical descriptors are capable of capturing, in a meaningful contraction, a great deal of relevant structural information. As illustrated in the case considered of imidazolidines, the graph theoretical descriptors are superior to traditional physicochemical descriptors, even when these are augmented with various quantum-chemically computed parameters.

8. New directions

It is hoped that we have succeeded in informing the reader of some aspects of the graph theoretical approach to structure–property and structure–activity problems. The graph theoretical approaches include, in addition to deriving correlations, the use of graph descriptors in other theoretical schemes, such as pattern recognition [30], the ranking of compounds, the search for substructure, and even the "design" of new compounds. All these methodologies have as a common part the use of various graph theoretical descriptors. A successful graph theoretical approach can be recognized as one which identifies the critical structural factor that dominates a property. In the case of chromatographic retention indices and boiling points, the discrimination of bond type, which is incorporated in the connectivity index, appears to be one such critical observation [31]. In the case of the aromaticity of conjugated hydrocarbons, it is the concept of conjugated circuits which plays the critical role [3]. In the case of isomeric variations in alkanes, p_2, p_3 coordinates suffice to reveal the regularity. All this indicates the combinatorial richness of molecular structure, but for an unsolved problem it need not be apparent which structural component is essential. The weighted path numbers, which can be reduced to molecular connectivity indices or alternatively to atomic ID numbers, appear to be *general* descriptors, particularly suited for comparison, similarity testing, searching for fragments, etc. Hence, it seems desirable to consider their extension to heteroatoms and to three-dimensional structures. Both of these important problems received recent attention and show promise. This is not the place to elaborate, but we will briefly comment on these most recent advances.

The "natural" way to discriminate heteroatoms has already been outlined for chlorine atoms of (arylimino)imidazolidines. With each heteroatom, one associates a characteristic (empirically determined) value for the corresponding diagonal matrix entry. There is in addition a "flexibility" associated with variations in off-diagonal entries of the adjacency matrix [32], fully analogous to similar approaches in extending the HMO method to π-systems with heteroatoms. It remains to be seen if there will be any connection between the empirically determined parameters based on selected structure–property correlations and those from quantum chemical calculations. If the answer is positive, one could be in a position to formulate "valency" rules, such as the well-known Slater rules for the construction of simple orbitals [33], or the rules of Kier and Hall for valency connectivity indices [34].

Three-dimensionality appears to be a more "difficult" problem, in part since there is no analogue in quantum chemical computations, which (unless a direct interaction of more distant centers is explicitly taken into account) are also devoid of three-dimensionality. The difficulty of the task can already be visualized for *cis* and *trans* butadiene, which in simple HMO theory are not discriminated. In order to discriminate between *cis* and *trans* butadiene, we have to take into account differences in their *geometry*. This immediately suggests the use of geometric matrices instead of adjacency matrices as a basis to represent molecules [35]. If we take the CC bond length as a unit, we obtain the geometric, or topographic, matrices shown in table 11. Observe that the matrices for *cis* and *trans* isomers differ; hence, if they are now viewed as "weighted" matrices, we can use the ALLPATH program [36] and derive the associated atomic path numbers, molecular path numbers, atomic ID and molecular ID numbers in the way that these quantities are derived from the adjacency matrix. Table 12 gives results for the two butadienes and compares them with the corresponding numbers derived from the adjacency matrix. As can be seen, the two sets of numbers are very comparable, the weighted path process being rather stable, i.e. no sudden variations are found among the descriptors for similar molecules or atoms in a similar environment. Thus, we are optimistic that path numbers (suitably weighted), whether arising from

Table 11

Adjacency matrix of butadiene graph compared with topographic (geometric) matrices for *cis* and *trans* butadienes (CC bond length assumed to be 1.00)

Butadiene graph	*cis*-butadiene	*trans*-butadiene
$\begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & \sqrt{3} & 2 \\ 1 & 0 & 1 & \sqrt{3} \\ \sqrt{3} & 1 & 0 & 1 \\ 2 & \sqrt{3} & 1 & 0 \end{pmatrix}$	$\begin{pmatrix} 0 & 1 & \sqrt{3} & \sqrt{7} \\ 1 & 0 & 1 & \sqrt{3} \\ \sqrt{3} & 1 & 0 & 1 \\ \sqrt{7} & \sqrt{3} & 1 & 0 \end{pmatrix}$

Table 12

The weigted path numbers $p(k)$, atomic ID numbers $a(k)$, and the molecular ID for butadiene, viewed as *cis*, *trans*, or ordinary graph

	Path numbers		
	$p(1)$	$p(2)$	$p(3)$
cis-butadiene	1.991	1.290	0.431
trans-butadiene	1.956	1.123	0.349
butadiene graph	1.914	0.707	0.250

	Atomic ID (path sums)		
	$a(1)$	$a(2)$	Molecular ID
cis-butadiene	2.916	2.796	7.7123
trans-butadiene	2.660	2.794	7.4276
butadiene graph	2.311	2.561	6.8713

molecular graphs, from "colored" graphs, i.e. graphs with heteroatoms, or, finally, from structures embedded in three-dimensional space, may offer a suitable characterization of chemical structure.

9. On the role of graph theory in chemistry

It appears appropriate to end this exposition on chemical structure from a graph theoretical perspective with some general comments on the role of graph theory in chemistry, and theoretical chemistry in particular. One should be reminded that, apparently, there is a prevailing impression among chemists that difficult problems lie ahead with regard to the computational difficulties arising from the proliferation of molecular integrals in quantum chemical computations when one strives for ever increased accuracy for ever larger molecular systems. Few, however, indicated that there are conceptual difficulties to be resolved as one extends the size of molecules and approaches biologically important systems. Primas [37] has been very explicit in his book in emphasizing the conceptual, not the computational, difficulties as the major obstacle to be resolved. Parr [38] also has raised questions concerning "extrapolation" of quantum chemistry to ever larger molecular systems, and so have several others. Consider some of the following questions:

(1) Why do some theoretical models give much better results than the underlying assumptions would justify?

(2) Why do some theoretical models apply outside the domain of their initial inauguration?

(3) Why do nonphysical quantities correlate with observables?

Illustrations of (1) include the Hückel MO model (and the underlying approximation of the nearest-neighbor interaction advanced by Bloch [39]), and the crystal field model, while an illustration of (2) includes the Hückel $4n + 2$ aromaticity rule, which can be mathematically justified for monocyclic rings only, yet it holds for catacondensed (but not many pericondensed) benzenoid hydrocarbons. Equally, we could here mention the Woodward–Hoffmann orbital symmetry rules [40], which operate even when symmetry is not an essential element of a system. Both (1) and (2) may apparently be viewed as problems that quantum chemistry may consider (and no doubt some have commented on them), but more "troublesome" is the case (3), since it clearly goes beyond quantum chemistry, which as a branch of quantum theory is confined only to observables. The list of commonly used non-observables is impressive, not perhaps by its length but by its content, which includes: potential energy (and, of course, potential surfaces), electrostatic potential, hybridization and hydrids, molecular orbitals, bond orders (both Coulson's as well as Pauling's), Kekulé valence structures (and, of course, conjugated circuits), resonance energy, bond dipoles, molecular surfaces, molecular volumes, etc.

Graph theory has already clarified some of these difficulties. For example, the Pauling bond order can be directly interpreted within the chemical graph theory and thus becomes a simple graph theoretical invariant or, if you will, a mathematical "property". Hence, when such bond orders are used to discuss observable CC bond lengths in a molecule, such as naphthalene or phenanthrene, one is merely considering a property–property correlation, where one property is a mathematical property (Pauling bond order) and the other is a physicochemical property (CC bond lengths). In such an approach, the dilemma of how to correlate a structure, which is not a number, with a property (commonly expressed as numbers) is resolved by considering a mathematical property instead of chemical structure as the reference object. Consider now a correlation using graph theoretical descriptors and a physicochemical property. Since physicochemical properties, at least in principle, follow from prescribed mathematical analyses (which is behind the axiomatics of quantum theory), we in fact see that one can view such correlations as mathematical property versus mathematical property correlations, where one class of mathematical properties are structural invariants and the other is frequently "substituted" with experimental properties in view of computational difficulties (or more correctly, computational inaccuracies of computed wave functions). The actual chemical properties that are measured by some instruments only serve to verify that our "advanced" mathematical model (quantum theory) applies, and that the particular approximations are adequate. The "other" mathematical models (graph invariants, structural invariants) only help us to visualize the results, since graph theoretical concepts are "transparent", while quantum chemical computations on larger systems are highly convoluted.

Hence, quantum chemistry plays the role of "perfect" model, while graph theory helps to "digest" very advanced computational models, so to speak. Graph theory can hopefully help one in building a visual image of highly elaborate models or to interpret partial results in terms of some underlying structural invariants. That is exactly what Coulson did when he introduced his bond order, which can be interpreted as a graph theoretical construction. However, that is not where graph theory ends; that is where it begins and where it makes "bridges" to quantum chemistry. Graph theory is basically concerned with consequences of the particular connectivity present in a system, rather than being interested in the origin of the connectivity, i.e. in "The Nature of Chemical Bonding" [41], it is interested in the "follow up" and, hence, operates at a lower "resolution" by accepting given bonding and continues to search for various consequences that the particular bonding implies. Hence, we can rightly refer to such concerns as the study of "The Nature of Chemical Structure".

Can one reconcile quantum chemistry with graph theory? Well, first of all there is no contradiction here. The two theoretical disciplines have different domains. They operate at different levels and are complementary rather than competitive or duplicative of one another. Some confusion about graph theory may have been caused by those who identify the Hückel MO as graph theory, when in fact the parallelism extends only to the mathematical equivalence between the two when considering HMO and graph spectral properties. Graph theory is a branch of mathematics, HMO is a chemical model, not mathematics, and only because of the *tacit* assumption that the interaction matrix (approximate Hamiltonian) in HMO theory is the adjacency matrix of the underlying graph do the two become computationally equivalent.

On the other hand, if we generalize graphs with variable weights, as has been illustrated in this paper, we can view general matrices as objects of generalized graph theory. Here, we only mention the geometric or topographic matrix (in which bonds are measured in CC bond-length units), besides the adjacency matrix. However, we can take a step forward and consider *any* matrix, including molecular Hamiltonian matrices, or matrices whose elements are selected molecular properties as objects of graph theoretical analysis. For example, to be specific we could take the Pariser–Parr–Pople matrix [42], associated with, say, anthracene, and consider it as a mathematical object for detailed analysis. Then, if one confines his or her interest to the spectral properties of such a matrix, one can claim, in the same spirit that people recognized HMO theory as graph theory, that the PPP method and the whole of traditional quantum chemistry is part of graph theory. However, the emphasis in graph theoretical studies is on structural invariants of such matrices, while the emphasis in quantum chemistry is on the particular structural invariants: eigenvalues and eigenvectors. It seems desirable, therefore, to maintain this distinction between quantum chemistry and graph theory, even though eigenvalues and eigenvectors are as much a part of graph theory as they are a part of quantum chemistry. The generalization of graphs to weighted graphs, which then allows one to take a step further and consider any matrix as a "weighted" graph, opens novel horizons for the chemical combinatorics and topology – which graph theory in essence is. However, let us again emphasize that our interst, from a graph theoretical position, remains with structural *invariants* as a tool for a better understanding of the nature of the chemical structure, i.e the description of a structure in terms of critical structural parameters. This supplements the interests of traditional quantum chemistry, concerned with eigenvalues and eigenvectors and their subsequent use as structural invariants in computing molecular properties.

Acknowledgements

The author would like to thank the referee for numerous valuable remarks which contributed to a more readable presentation of the subject. Thanks are also due to Professor D.J. Klein for examining the manuscript and suggesting improvements in form and content of a number of passages in the text.

References

[1] *The New Lexicon Webster's Dictionary of the English Language* (Lexicon Publications, New York, 1988 edition).
[2] N.J. Turro, Angew. Chem. Int. Ed. Engl. 25(1986)882.
[3] M. Randić, Chem. Phys. Lett. 38(1976)69; M. Randić, J. Amer. Chem. Soc. 99(1977)444; M. Randić, Int. J. Quant. Chem.: Quant. Biol. Symp. 11(1984)137.
[4] M. Randić, J. Chem. Inf. Comput. Sci. 24(1984)164.
[5] H. Hosoya, Bull. Chem. Soc. Japan 44(1971)2332.
[6] M. Randić, J. Amer. Chem. Soc. 97(1975)6609.
[7] H.F. Hameka, J. Chem. Phys. 34(1961)1966.
[8] P.S. O'Sullivan and H.F. Hameka, J. Amer. Chem. Soc. 92(1970)25;
L.V. Haley and H.F. Hameka, J. Amer. Chem. Soc. 96(1974)2020.
[9] M. Randić, Chem. Phys. Lett. 53(1978)602.
[10] M. Randić and C.L. Wilkins, Chem. Phys. Lett. 63(1979)332.
[11] M. Randić and C.L. Wilkins, J. Phys. Chem. (E. Bright Wilson Festschrift).
[12] M. Randić and N. Trinajstić, MATCH.
[13] M. Randić, J. Magn. Res. 39(1980)431.
[14] M. Randić and N. Trinajstić, Theor. Chim. Acta 73(1988)233.
[15] Y. Miyashita, T. Okuyama, H. Ohsako and S.-i. Sasaki, J. Amer. Chem. Soc. 111(1989)3469.
[16] R. Langebach, C. Kruszynski, R. Gingell, T. Lawson, D. Nagel, P. Pour and S.C. Nesnow, in: *Structure–Activity Correlation as a Predictive Tool in Toxicology*, ed. L. Goldberg (McGraw-Hill, New York, 1983), Ch. 16.

[17] M. Randić, B. Jerman-Blažič, D.H. Rouvray, P.G. Seybold and S.C. Grossman, Int. J. Quant. Chem.: Quant. Biol. Symp. 14(1987)245.
[18] D. Lednicer and L.A. Mitscher, *The Organic Chemistry of Drug Synthesis*, Vol. 1 (Wiley-Interscience, New York), pp. 286–293.
[19] G. Klopman and H.S. Rosenkranz, Mutation Res. 126(1984)227.
[20] M. Randić, a chapter in the forthcoming book *Similarity in Chemistry*, ed. M. Johnson and G.M. Maggiora (Wiley, New York), in press.
[21] R.P. Sheridan and R. Venkataraghavan, Acc. Chem. Res. 20(1987)322;
R.E. Carhart, D.H. Smith and R. Venkataraghavan, J. Chem. Inf. Comput. Sci. 25(1985)64.
[22] C. Hansch, Acc. Chem. Res. 2(1977)120.
[23] N. Trinajstić, D.J. Klein and M. Randić, Int. J. Quant. Chem.: Quant. Biol. Symp. 20(1986)699;
N. Trinajstić, M. Randić and D.J. Klein, Acta Pharm. Yugosl. 36(1986)267.
[24] N. Trinajstić, *Chemical Graph Theory*, Vols.1 and 2 (CRC Press, Boca Raton, FL, 1985).
[25] B.M.W.M. Timmermans and P.A. van Zwieten, J. Med. Chem. 20(1977)1636.
[26] M. Randić, submitted for publication.
[27] L.B. Kier and L.H. Hall, *Molecular Connectivity in Chemistry and Drug Research* (Academic Press, New York, 1976).
[28] A.A. Balandin, Uspekhi Khim. 9(1940)390.
[29] L. Spialter, J. Chem. Doc. 4(1964)261.
[30] T. Okuyama, Y. Miyashita, S. Kanaya, H. Katsumi, S.-i. Sasaki and M. Randić, J. Comput. Chem. 9(1988)636.
[31] M. Randić, J. Chromatogr. 161(1978)1.
[32] S.C. Grossman, B. Jerman-Blažič and M. Randić, Int. J. Quant. Chem.: Quant. Biol. Symp. 12(1986)123.
[33] J.C. Slater, Phys. Rev. 36(1930)57.
[34] L.B. Kier and L.H. Hall, J. Pharm. Sci. 65(1976).
[35] M. Randić, Studies in Phys. Theor. Chem. 54(1988)101; M. Randić, Int. J. Quant. Chem.: Quant. Biol. Symp. 15(1988)201.
[36] M. Randić, G.M. Brissey, R.B. Spencer and C.L. Wilkins, Computers and Chem. 3(1979)5.
[37] H. Primas, *Classical Observables in Molecular Quantum Mechanics.*
[38] R.G. Parr, Proc. Natl. Acad. Sci.,
[39] F. Bloch, Z. Phys. (1928).
[40] R.B. Woodward and R. Hoffmann, Angew. Chem. 81(1969)797.
[41] L. Pauling, *The Nature of the Chemical Bond* (Cornell University Press, Ithaca, NY, 1940).
[42] R. Pariser and R.G. Parr, J. Chem. Phys. 21(1953)466;
J.A. Pople, Trans. Faraday Soc. 49(1953)1375.