Journal of Mathematical Chemistry 5(1990)43–72

COMPARABILITY GRAPHS AND MOLECULAR PROPERTIES: IV. GENERALIZATIONS AND APPLICATIONS*

D. BONCHEV, V. KAMENSKA and O. MEKENYAN
Higher School of Chemical Technology, Burgas 8010, Bulgaria

Received 9 March 1989

Dedicated to the memory of Professor Oskar E. Polansky, a pioneer of chemical graph theory

Abstract

The development of a recently proposed method for calculating molecular properties is outlined. The approach is based on the idea of constructing optimized compound samples for structure-property or structure-activity correlations by means of the so-called comparability graphs (CG) of isomeric compounds. A dynamic comparability principle is devised, proceeding from a series of standard molecular rearrangements described in graph-theoretical terms as rules on molecular branching and cyclicity. An extension of the approach is presented for both the construction of CG's and their combination for variable numbers of atoms. The method is applied to various physico-chemical properties, which are thus divided into three groups according to the degree to which they are conditioned by molecular topology. The Wiener topological index is shown to produce a highly linear correlation with the alkane critical densities and volumes, as well as with their heats and entropies of vaporization.

1. Introduction

More than 300,000 chemical compounds are synthesized every year in chemical laboratories. The examination of these compounds to reveal their practical use usually requires sophisticated intellectual efforts and considerable financial resources. A typical example is the search for new drugs, where only one out of four thousand compounds selectively synthesized proves to be of real importance and only one out of fifteen thousand comes on the market. Related to this, the emphasis is more and more shifted to the syntheses of such compounds where properties are evaluated in advance. A new branch of chemical science termed "molecular engineering" is now in progress, which combines methods for designing new compounds, mixtures, composites, etc. with the evaluation of their properties by means of quantitative structure-property (QSPR) [4] and structure-activity (QSAR) [5,6] relationships. Mathematical chemistry [7–10] and,

*See refs. [1–3] for Parts 1 to 3, respectively, of this series.

© J.C. Baltzer AG, Scientific Publishing Company

mostly, chemical graph theory [11–18] is in use in converting chemical structure into a number termed molecular descriptor or structural index, including here topological [19–23], information-theoretic [24–31] indices, etc.

Looking more closely at structure-property and structure-activity correlations, a number of questions arise, the most fundamental being *how to present molecular structure?* and *how to construct the correlation sample of compounds?* The search for the most convenient molecular descriptors is still an open question in mathematical chemistry, the number of structural indices proposed becoming tremendous. Less attention is paid, however, to the proper selection of compounds for the needs of correlation analysis. The straightforward attempts to approach the problem by including the whole group of compounds under consideration are unacceptable, both for theoretical and practical reasons. Some of the latter are illustrated in fig. 1, where the property X is

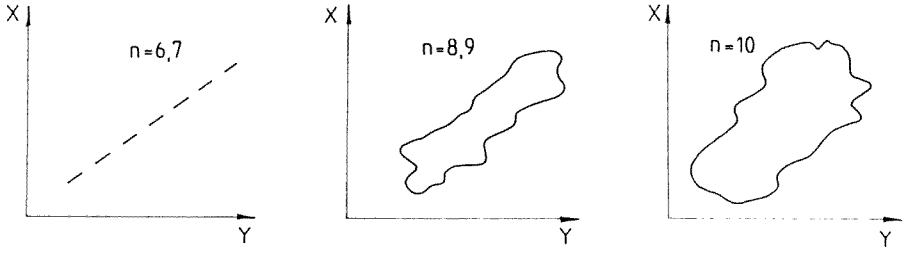


Fig. 1. Trends in correlating the molecular property X with the molecular descriptor Y at an increased number of carbon atoms n for a homologous series of compounds.

correlated with the molecular descriptor Y for a homologous series of compounds. This is a situation frequently faced: the trend in the X-Y correlation, albeit pronounced for a smaller number of carbon atoms, becomes more and more diffuse with their increase.

Figure 1 indicates clearly the need of specific selection of the compound correlation samples. The principles of such a selection might be various. Structural similarity is one of the approaches, which is perhaps realized best within the DARC-PELCO method [32–34]. An alternative way is the preliminary ordering of compounds making use of specific structural criteria [35–38]. Differing from these considerations of chemical structure in its statics, we recently developed [1–3] a dynamic principle of compound ordering based on molecular topology and, more specifically, on molecular branching and cyclicity. The idea was realized by means of the so-called *comparability graphs* constructed for isomeric compounds. The present study is meant to re-examine this approach and extend it to variable numbers of atoms by combining different comparability graphs. Its importance for the calculation of molecular properties will be studied in detail, including also the derivation of optimized structure-property correlations.

2. Comparability graph and its construction

In building up the comparability graph for a group of isomeric compounds, one needs a system of topological rules (rules on modeling branching, cyclicity, etc.). Each of these rules describes a specific molecular rearrangement which is associated with a regular alteration of a certain graph-invariant. Such rules have been derived for molecular branching [39,40] and cyclicity [41–45] proceeding from the Wiener number [46] which represents the total distance of the molecular graph. Recently, some of these branching rules have also been derived by means of the so-called graph connections [47].

Once the system of topological rules is available, it is applied to the set of isomeric compounds under consideration. A partial ordering of these compounds results since each molecular rearrangement regarded increases the skeleton branching or cyclicity, as follows from the respective decrease in the Wiener number. This partial ordering is expressed by simple directed graphs with vertices denoting the compounds that take part in the respective molecular transformations. Then, these graphs are superimposed, thus arriving at the comparability graph (CG) of the whole series of compounds. Each vertex in this graph is connected with all those which represent the compounds the respective isomer can be converted into or can be obtained from the standard transformations described by the topological rules.

It should be noted that the comparability graph is built proceeding from the topological index by Wiener, which is the only graph-invariant used to characterize molecular branching and cyclicity by a detailed system of rigorously derived rules. Being an inverse proportional measure of molecule compactness, this index proved to be of considerable importance in modeling the properties of molecules, polymers, and crystals [48]. The physical background of the Wiener number's wide applicability has very recently been revealed [31] through its relation to another graph-invariant, the number of self-returning walks SRW which, in its turn, is related to the quantum-mechanical method of moments [49]. More specifically, it has been shown that the most typical branching rules are associated with an increase in the number of self-returning walks of degree 4 ($SRW^4 > 0$), while three out of the ten branching rules do not produce any change in this graph-invariant ($SRW^4 = 0$).

We thus arrive at the idea of modifying the Wiener-based comparability graph CG(W) by comparing it to another (pseudo) comparability graph CG(TI_i) based on a different graph-invariant TI_i. The term "pseudo" recalls that CS(TI_i)'s are built by using numerical values but not rigorously derived topological rules. In addition to SRW^4, the molecular connectivity index χ by Randić [50] and non-adjacency number Z by Hosoya [51] have also been taken as representative graph-invariants. Other graph-invariants, such as the Randić molecular path code [52], the Bonchev and Trinajstić information-theoretic indices I_D^W and I_D^E [39], the information-theoretic analogue of the Hosoya index I_{pc} [39], and the largest eigenvalue of the characteristic polynomial x_i [53], have been previously shown [2,3] to produce no further changes, as compared with the basic four indices.

The final stage of the CG construction thus begins with the building up of CG(SRW^4), CG(χ), and CG(Z). Then, the three CG's are combined into a single connecting comparability graph CG(SRW^4, χ, Z). In the superimposing procedure, two isomers X and Y are regarded as *non-comparable* and hence they cannot be located on the same directed path if the relationships

$$X(\mathrm{TI}_i) < Y(\mathrm{TI}_i),$$
$$X(\mathrm{TI}_j) \geq Y(\mathrm{TI}_j) \tag{1}$$

holds simultaneously. In this case, X and Y belong to different graph paths.

Finally, CG(W) and CG(SRW^4, χ, Z) are compared, eliminating those arcs XY in CG(W) for which

$$X(W) < Y(W), \tag{2}$$

$$X(SRW^4, \chi, Z) \geq Y(SRW^4, \chi, Z) \tag{3}$$

hold simultaneously.

The ultimate comparability graph CG(W, SRW^4, χ, Z), contains along its path only *comparable* isomers; these are isomers which are characterized by a regular change in the four selected topological indices ($\Delta W < 0$, $\Delta SRW^4 > 0$, $\Delta\chi < 0$, $\Delta Z < 0$) upon the consecutive molecular rearrangements occurring along each directed path in CG. It is the basic assumption of the comparability graph concept that this dynamic principle of ordering molecular structures along each of the CG paths could reflect the regularities in molecular data and provide *optimized* correlation samples of isomeric compounds. A large range of experimental evidence is presented in the remainder of this work in support of our assumption.

In evaluating the comparability graph concept, one may judge that its significance largely depends on the completeness of the scheme presented. This implies a more general formulation of molecular branching [54] and cyclicity rules for a more complete description of the isomer interconversions, as well as a proper choice of the representative graph-invariants used for isomer ordering. We suppose that the selection of topological indices made in our approach is a sufficiently representative one because it is based on quite different topological features of molecules: graph distances, connectivity, and non-adjacency. On the other hand, the use of CG combining several graph-invariants is presumably preferable to the separate use of several CG's, each one based on a single invariant. Yet, the opposite point of view, regarding a certain molecular property to be described best by a sole topological index, also deserves attention. This will be examined in the next publication of this series [55].

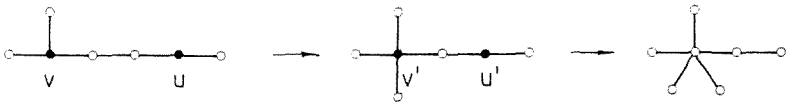
3. Molecular branching rules

The procedure of constructing CG and its relevance to molecular properties is illustrated in detail in the following section for the series of alkane compounds. To

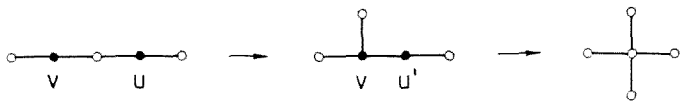
enable these considerations, we briefly outline the ten rules on molecular branching previously derived [39,2,3]. They can be formulated in a unified manner:

In acyclic graphs, the branching increases (the Wiener number decreases) when, for a constant number of vertices,

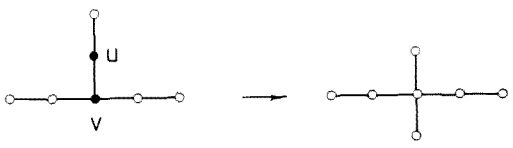
Rule 1. . . . the number of branches attached to a certain vertex increases at the expense of chain shortening.



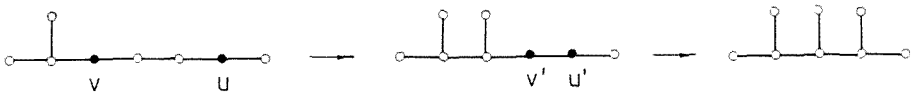
Rule 2. . . . a chain graph transforms into a tree graph and the latter converts further into a star graph.



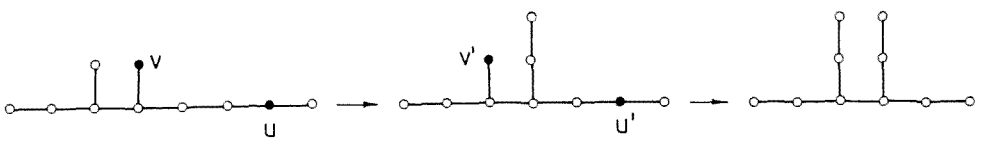
Rule 3. . . . the number of branches attached to a certain vertex increases at the expense of shortening a long branch connected to the same vertex.



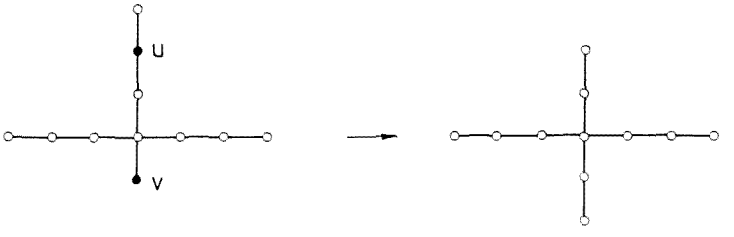
Rule 4. . . . the number of branches attached to different vertices increases at the expense of chain shortening.



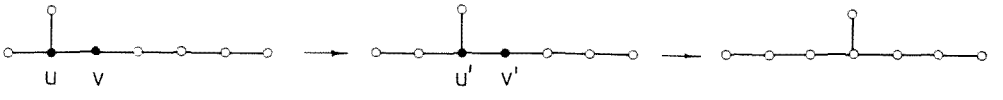
Rule 5. . . . the length of branches increases at the expense of chain shortening.



Rule 6. . . . the length of a branch increases at the expense of shortening a longer branch attached to the same chain vertex.



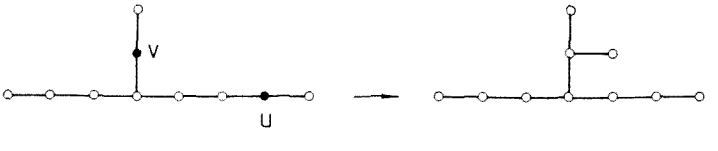
Rule 7. . . . a branch displaces from the terminal to a more central chain vertex.



Rule 8. . . . and branches a branch displaced from a chain vertex to another one having an equivalent position along the chain.

Rule 9. . . . the number of branches attached to a certain vertex increases at the expense of shortening a long branch connected to the nearest neighboring vertex.

Rule 10. . . . a branched branch is formed at the expense of chain shortening.

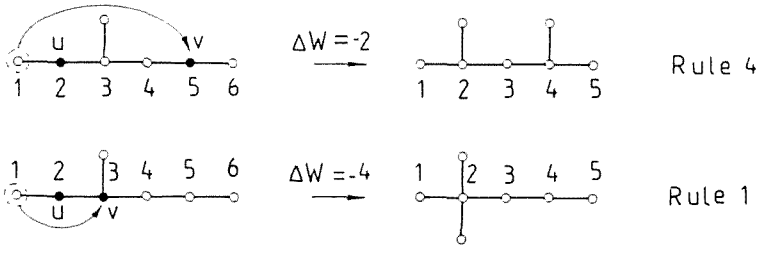


As seen from the illustrations, rules 1 to 10 deal with vertex transfers that alter some of the factors presumably affecting molecular branching: the number, length, and position of branches, the number of branches to a certain vertex, etc. All these are associated with a decrease in the Wiener number. Transfers of terminal vertices (including branch ones) between two other non-equivalent graph vertices are mainly treated. Transfers of fragments having more than one atom can also be handled in this way since, as was shown recently [56], the change in the Wiener number sign does not depend on the number of atoms in the transferred fragment n_{11}:

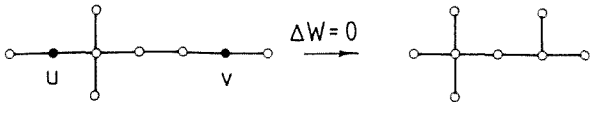
$$W = n_{11}(d_u - d_v), \tag{4}$$

where d_u and d_v are the distances (or distance numbers) of the vertices u and v, the latter being the nearest neighbors to the transferred fragment in the initial and final graph, respectively.

Another comment on molecular branching rules is related to the term "chain shortening" used in their definitions. In our previous studies, we disregarded part of the chain shortenings for which, in addition to new branch formation, an opposing structural factor takes place. As shown below, this factor is the central-to-terminal displacement of the branches located closer to the shortened chain end (the inverse of rule 7).



Detailed analysis, however, indicates that the formation of a new branch usually strongly dominates over the second factor, and a systematic decrease in the Wiener number occurs. With main chain elongation, a few cases appear with $\Delta W = 0$ or even $\Delta W > 0$. Instead of eliminating all "short side" chain shortenings, in this study we discard solely the latter cases, regarding them as examples for the limits beyond which the Wiener number cannot be used to characterize molecular branching. As a matter of fact, in these cases the Wiener index is not in a state to reflect the increase in molecular branching, due to the formation of a new branch.



4. Examples of comparability graph construction

The procedure of constructing CG is illustrated for the series of nine isomeric C_7-alkanes (fig. 2). In fig. 3, the comparability graph CG(W) of these isomers is shown, the branching rules used being denoted along each edge symbolizing an isomer conversion (see ref. [2] for more details). Each isomer located on a certain path in CG(W) is characterized by a Wiener number smaller than those of the isomers preceding it. For example, for the ABHI path, $W(A) < W(B) < W(H) < W(I)$ holds.

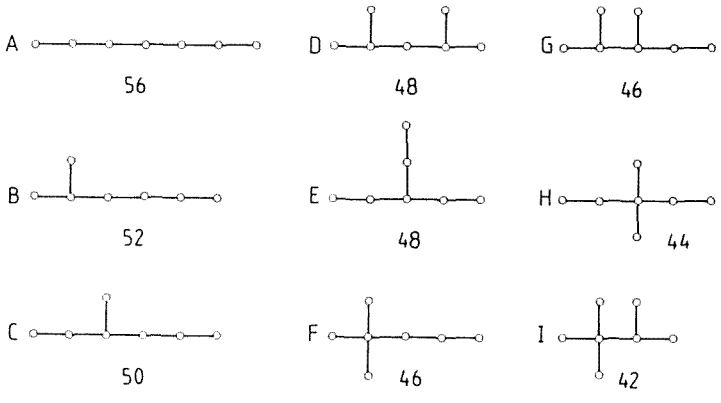


Fig. 2. Hydrogen-depleted molecular graph of C_7-alkanes and their Wiener numbers.

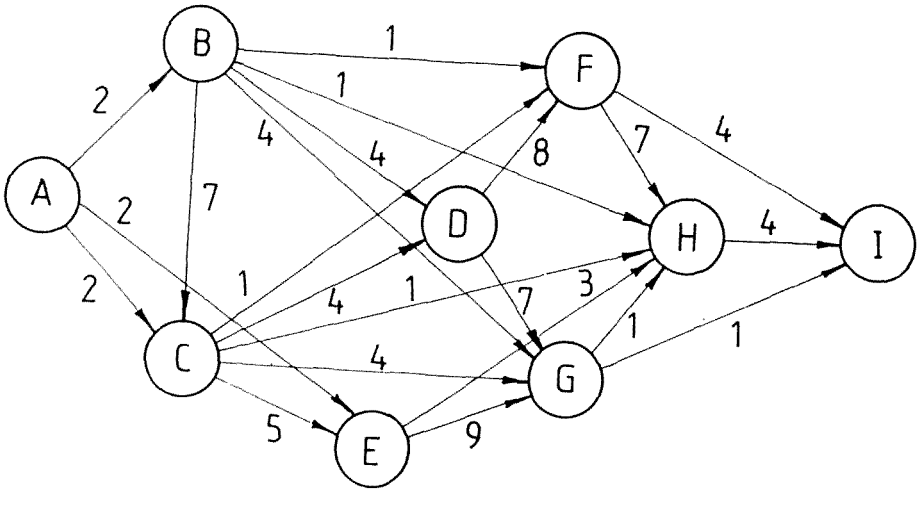


Fig. 3. The Wiener number-based comparability graph of C_7-alkane isomers CG(W). The letters correspond to those in fig. 2.

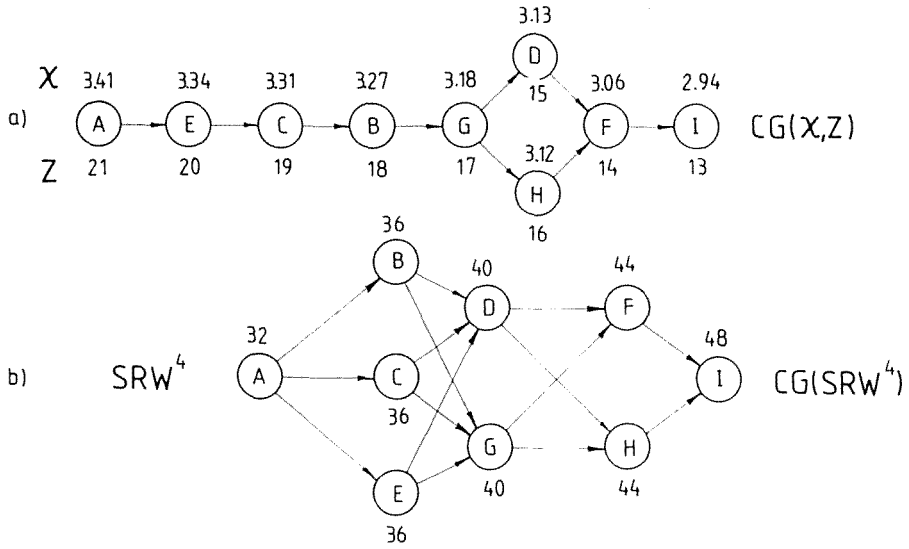


Fig. 4. Pseudo-comparability graphs of C_7-alkanes which are not based on molecular branching rules but on the numerical values of: (a) the Randić (χ) and Hosoya (Z) indices; (b) the number of self-returning walks of degree 4 (SRW^4).

In fig. 4, the pseudo-comparability graphs of the same series of C_7-alkanes are shown, as obtained from the numerical values of the topological indices of Randić and Hosoya, as well as from the number of self-returning walks of degree 4. As seen, the first two indices produce almost the same isomer ordering, the only difference being in interchanging isomers D and H. SRW^4 specifies three groups (a triplet and two doublets) of non-comparable isomers which cannot be located within the same graph path.

Comparison of CG(W) with CG(χ, Z) and CG(SRW^4) proceeding from relationships (2) and (3) results in the ultimate comparability graph of C_7-alkanes CG(W, χ, Z, SRW^4), which is shown in fig. 5. Only four out of the twenty-two arcs contained in CG(W) are eliminated in this procedure. These are all graph transformations obeying molecular branching rules 5 ($C \rightleftarrows E$) and 7 ($B \rightleftarrows C$, $D \rightleftarrows G$, and $F \rightleftarrows H$) for which χ, Z, and SRW^4 either reverse the ordering produced by W or cannot distinguish between the respective pairs of isomers.

The comparability graph shown in fig. 5 is much more complete than that reported in our preceding study [2]. It incorporates eighteen instead of thirteen isomer rearrangements. The inclusion of three of them was advocated in the foregoing: two transfers of a terminal vertex taken from the shorter side of the graph main chain ($C \rightarrow D$ and $C \rightarrow F$) and a joint transfer of two terminal vertices ($B \rightarrow H$). In addition, the $B \rightarrow F$ transition is included, which was eliminated in ref. [2] due to the presence of a longer path $B \rightarrow D \rightarrow F$ between the same vertices. Such direct rearrangements ought to be preserved in order to prevent the complete path disconnecting in cases when

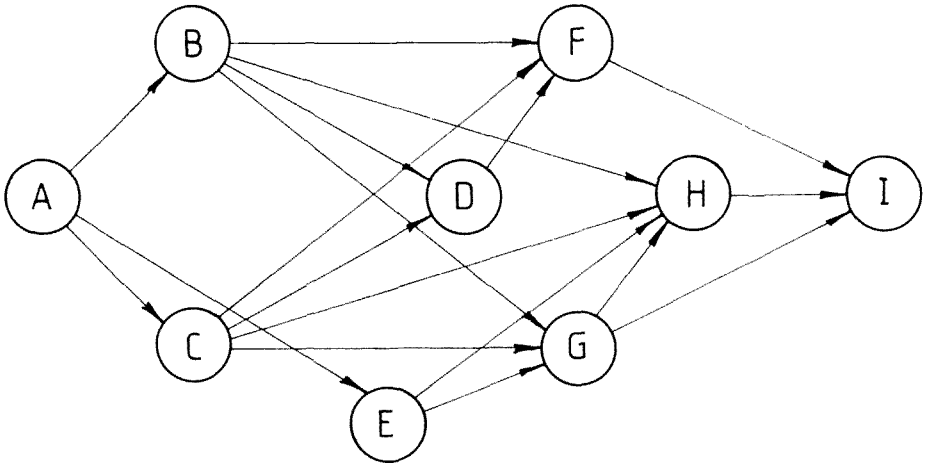


Fig. 5. Ultimate C_7-alkanes comparability graph CG(W, χ, Z, SRW^4)

one of the indirect transformations (such as B → D and D → F) is eliminated upon the CG or CG-subgraph construction. Finally, the G → H transformation is also preserved, which has to be associated with branching rule 1 rather than with rule 7, as was done in ref. [2].

The incorporation of five new isomer transformations in the C_7-alkanes comparability graph also results in a larger number of paths ordering isomeric molecules: 5 paths of length five (ABDFI, ABGHI, ACDFI, ACGHI, and AEGHI) and 8 paths of length four (ABFI, ABGI, ABHI, ACFI, ACGI, ACHI, AEGI, AEHI) versus only 1 path of length five and 5 paths of length four in our previous study. All these paths contain comparable isomers, i.e. isomers which can be consecutively transformed into the next one along the sequence. This isomer transformability is regarded as a crucial condition in selecting optimized correlation samples of compounds.

In dealing with C_8-alkanes (fig. 6), the same procedure produces a CG(W) graph having eighty edges. Twenty-one of them are then eliminated by conditions (2) and (3) after comparison with the two pseudo-comparability graphs CG(χ, Z) and CG(SRW^4). The ultimate C8-alkanes comparability graph CG(W, χ, Z, SRW^4) which thus results is shown in table 1 by its connection table. It incorporates much more isomer interconversions than the graph reported earlier [2] (59 versus 36). Eight of the new transformations (BF, CL, DL, GL, JQ, JR, KQ, KR) have been produced in ref. [2], but were not taken into account due to the presence of two-edge paths between the same pairs of vertices Four other cases deal with a transfer of a terminal vertex taken from the shorter side of the main chain (CE, CF, LQ, LN). As a consequence of eq. (4), transfers of fragments containing 2, 3, and 4 vertices were also included, which generated eight more molecular rearrangements (BL, KP; EO, ER, IR; and EN, IQ, NS, respectively). Finally, three other transformations needed a slight modification of some rules: GI (rule 4′), IL (rule 8′), MQ (rule 9′). As a whole, the CG presented in table 1 contains 56 paths of

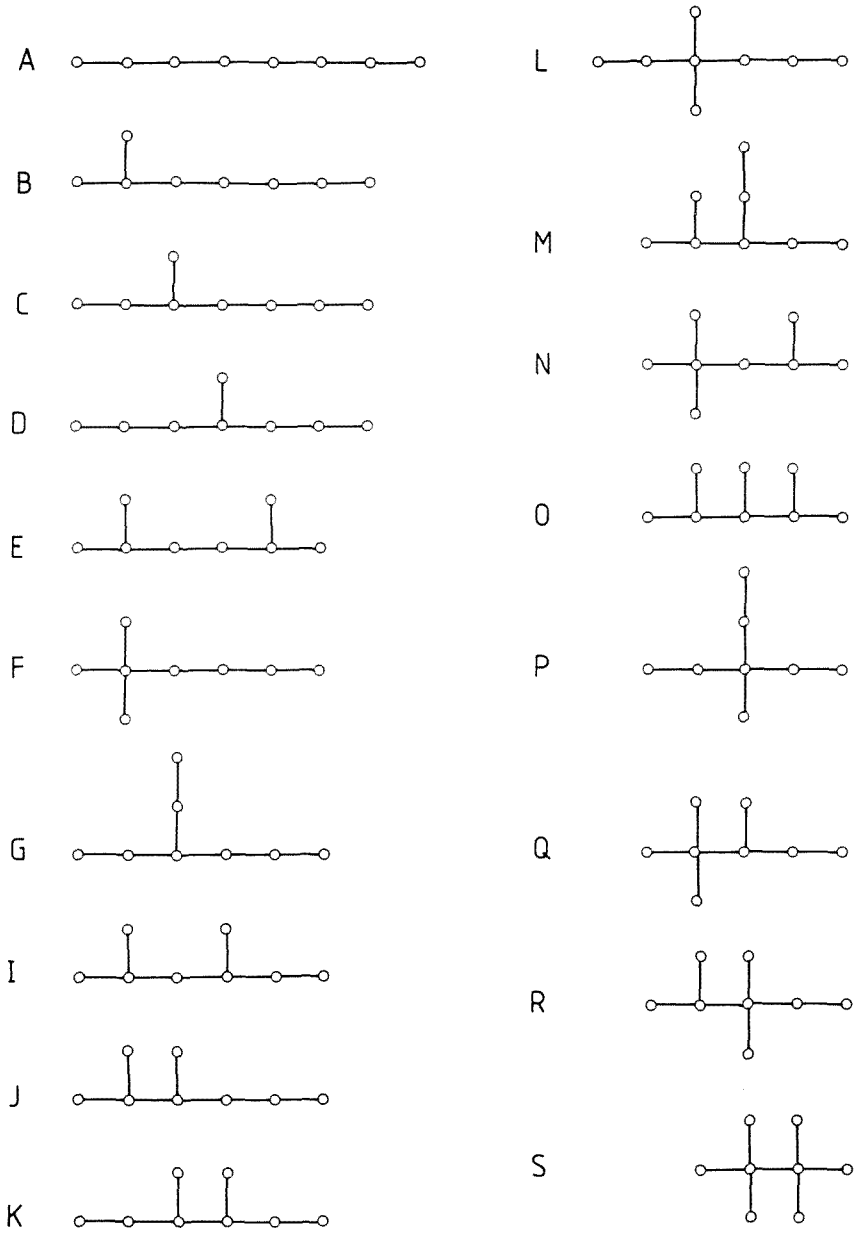


Fig. 6. Hydrogen-depleted molecular graphs of C_8-alkanes.

length six and 48 paths of length five versus 25 and 3 such paths, respectively, from ref. [2]. This considerably improves the basis for structure-property correlations.

Inspection of the eliminated graph transformations indicated that again these deal with branching rules 7 (fifteen cases) and 5 (five cases). The latter describe a branch

Table 1

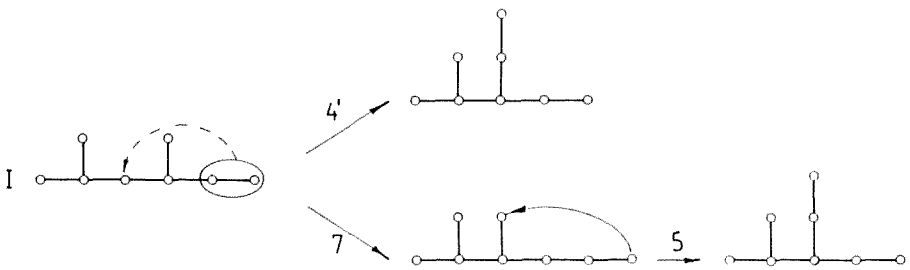
A connection table for the ultimate comparability graph of C_8-alkanes shown in fig. 6

A	B, C, D; G*
B	E, F, I, J; L, M
C	E, F, I, J, K, L; P
D	I, J, K, L, M
E	F; ; O, R; N
F	N, Q
G	I, J, K, L, M, P
I	L, O, N; ; R; Q
J	L, O, Q, R
K	L, O, Q, R; P
L	Q, R, N
M	O, Q
N	; ; ; S
O	Q, R
P	R
Q	S
R	S

*The semicolons separate the cases referring to transfer of fragments of 1, 2, 3, and 4 vertices, respectively.

displacement toward a more central position in the chain and a branch elongation at the expense of chain shortening:

The last eliminated transformation (IM), which was initially interpreted as a transfer of two terminal vertices to form a new branch to the chain (a modified rule 4), may be decomposed into two elementary steps which are described by the same rules 7 and 5. This indicates that transfers of fragments having more than one atom shoud be considered only when such a fragment is attached to a main chain vertex of degree 2.



5. Comparability graphs and regularities in molecular properties

Representing the dynamics of molecular topology by means of comparability graphs, one must next investigate the degree to which they reflect the regularities in compound properties. Arriving at a high degree of correspondence would be worthwhile, both from a theoretical and a practical viewpoint.

With this aim in mind, we have examined a large number of thermodynamic and other physico-chemical properties of C_7-, C_8-, and C_9-alkanes. Each of these properties was checked in order to find whether the general trend of increasing or decreasing in value, when going from the linear isomer to the most branched one, is followed by each of the molecular rearrangements given by the CG edges. The resulting statistics is presented in table 2. The three numbers in each of its columns denote the number of cases in which the property under consideration changes in the same direction, does not change, and changes in a direction opposite to that of the CG paths, respectively. In several cases, the sum of these three numbers is smaller than the total number of respective CG edges due to the lack of experimental data for one of the isomers. Data have not been found either for six C_9 properties.

The most essential result from an inspection of table 2 is that the examined twenty-seven alkane properties are roughly divided into three groups. The first one qualifies nine properties as strongly dependent on molecular topology or, more precisely, on molecular branching as characterized by the four topological indices. These are entropy of vaporization (ΔS_v), the Pitzer eccentric factor (Ω), coefficients C and A in the Antoine equation for the vapor pressure, octance number (N_0), entropy of formation (ΔS_f^0), standard entropy in gas and liquid state (S_g^0 and S_l^0), and enthalpy for the ideal gas state ($H_0 - H_0^0$). All these deviate from the CG path trends in less than 7% of the cases. The number of deviations is considerably larger (up to 25%) for the second group of seven properties. However, they may be qualified as branching dependent. These are parachor (a surface tension and density function, Ph), heats of vaporization (ΔH_v), the critical pressures, volumes, and densities (p_{cr}, V_{cr}, and ρ_{cr}, respectively), boiling points (T_B), and gas chromatographic retention indices (RI). The remaining eleven properties are almost independent of molecular branching since their deviations from the CG path trends span 25% to 50% of all cases. These are the coefficient B in the Antoine equation for the vapor pressure, the refractive index (n_d^{20}), relative density (d_4^{20}), molecular volume (V_M), molecular refraction (R_M), surface tension (σ), heats of formation and combustion in both gas and liquid state ($\Delta H_{f,g}^0$, $\Delta H_{f,l}^0$, $\Delta H_{c,g}^0$, and $\Delta H_{c,l}^0$), and critical temperatures (T_{cr}). Actually, no data were available for the heats of combustion in gas and liquid state, but their high similarity in behaviour with the respective heats of formation of C_7- and C_8-alkanes is a good reason to expect such a similarity for C_9 cases as well. Clearly, the possibilities for calculating the properties from the third group, proceeding from optimized isomer samples as produced by the CG's, are strongly limited. This point will be further discussed in this section by making use of the CG subgraphs constructed for each one of the properties under consideration.

Table 2
Regularities and irregularities in alkane properties [57] as produced by the comparability graphs of C_7-, C_8-, and C_9-alkanes

Properties \ Alkanes			C_7	C_8	C_9
N	Number of CG edges		18	59	99
1	ΔS_v		18, 0, 0	56, 0, 3	98, 0, 1
2	Ω	[58]	18, 0, 0	58, 1, 0	–
3	C		18, 0, 0	56, 0, 0	97, 0, 2
4	N_0	[59]	14, 0, 1 (15)	59, 0, 0	–
5	A		18, 0, 0	55, 1, 0	92, 0, 7
6	ΔS_f^0	[60]	17, 0, 1	58, 0, 1	–
7	S_g^0		17, 0, 1	56, 0, 3	96, 0, 3
8	S_l^0		12, 0, 1 (13)	–	93, 0, 6
9	$H^0 - H_0^0$		17, 0, 1	53, 0, 6	93, 0, 6
10	Ph	[58]	18, 0, 0	50, 0, 6 (56)	–
11	ΔH_v		18, 0, 0	56, 0, 3	84, 2, 13
12	p_{cr}		16, 0, 2	52, 1, 6	85, 10, 4
13	V_{cr}		13, 0, 5	44, 2, 13	86, 0, 13
14	ρ_{cr}		13, 0, 5	43, 2, 14	83, 5, 11
15	T_B		17, 0, 1	48, 0, 11	80, 0 19
16	RI	[61]	16, 0, 2	47, 1, 11	72, 0, 27
17	B		16, 0, 2	41, 0, 18	71, 0, 28
18	n_d^{20}		10, 0, 8	36, 0, 20 (56)	74, 0, 25
19	V_m		10, 0, 8	33, 0, 23 (56)	73, 0, 26
20	d_4		10, 0, 8	33, 0, 23 (56)	70, 0, 29
21	R_m		10, 0, 8	35, 0, 21 (56)	68, 1, 30
22	σ		13, 0, 5	33, 0, 23 (56)	56, 5, 38
23	$\Delta H_{c,g}^0$		16, 0, 2	49, 0, 10	–
24	$\Delta H_{c,l}^0$		16, 0, 2	45, 0, 14	–
25	$\Delta H_{f,g}^0$		16, 0, 2	49, 0, 10	67, 1, 31
26	$\Delta H_{f,l}^0$		16, 0, 2	44, 0, 15	57, 0, 42
27	T_{cr}		11, 0, 7	30, 0, 29	53, 0, 46

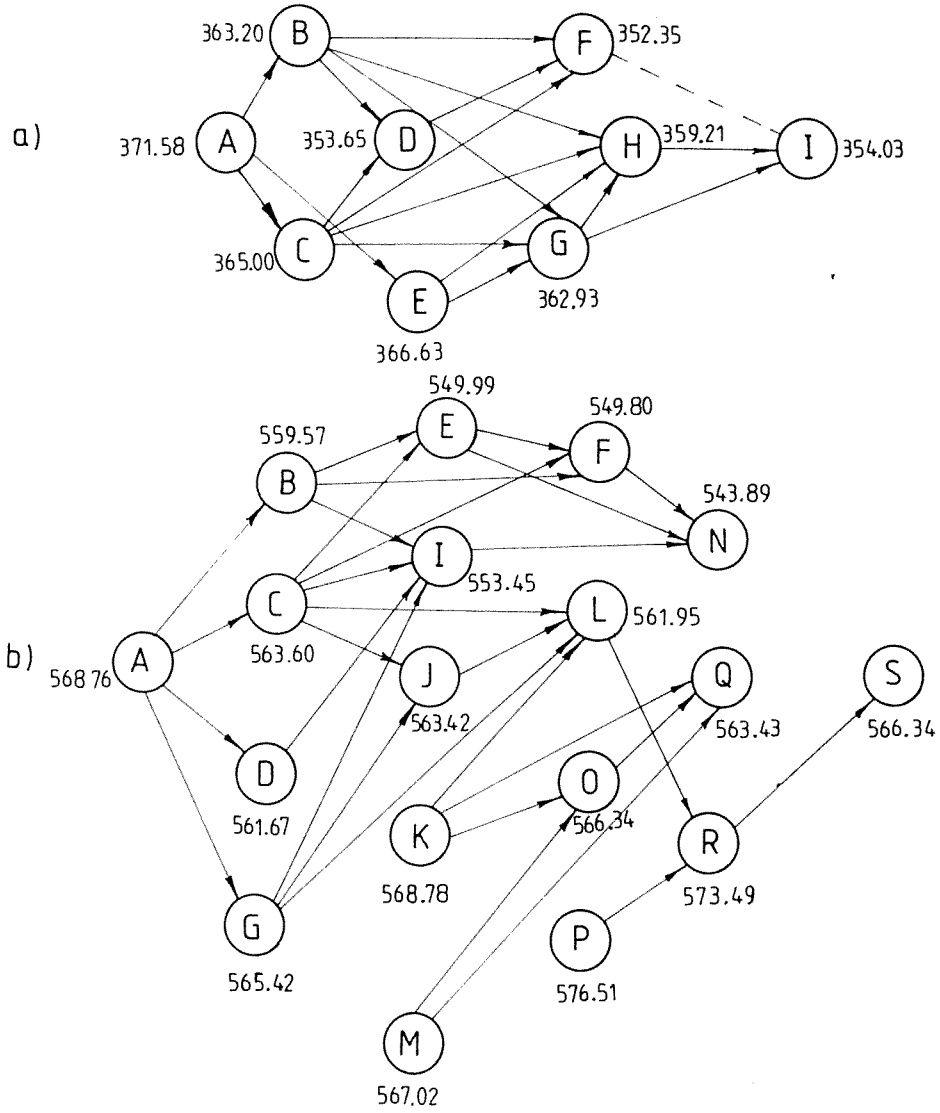


Fig. 7. (a) The boiling point subgraph of the C_7-alkanes CG; (b) the critical temperature subgraph of the C_8-alkanes CG (both temperatures in K) [57].

This idea is illustrated in figs. 7a and 7b. Figure 7a represents the boiling point subgraph of the C_7-alkanes CG. The dashed line is the only missing graph edge as compared to the C_7-alkanes CG. Otherwise, the ordering of isomer F before isomer I is the single discrepancy between the orderings of the C_7-alkanes according to their comparability graph and boiling point values.

Figure 7b exemplifies the CG-subgraph concept for C_8-alkane critical temperatures, the molecular property that most strongly disobeys the ordering produced by the

comparability graph. Only thirty out of fifty-nine isomer pairwise orderings are preserved in this case. As a result of this, the number and size of the optimized correlation samples of isomers is strongly reduced. Thus, the 56 graph paths containing six isomers and 48 paths with five isomers in each path are reduced to 2, 6, 7, 3, and 2 paths incorporating 6, 5, 4, 3, and 2 isomers, respectively. This makes it impossible to devise statistically significant structure-property correlations for the critical temperature of a group of isomeric alkanes. However, the strongly restricted regularity in T_{cr} and the other ten properties of the third group may be utilized in the search for structure-property correlations when dealing with compounds with a variable number of carbon atoms. This problem will be discussed in one of the sections to follow.

6. Comparability graphs of benzenoid hydrocarbons

Although the comparability graph methodology is presented here in detail for acyclic hydrocarbons, it is applicable to cyclic structures as well. In part I of this series [1], we built up the CG's for cata-fused benzenoids having 4, 5, and 6 rings, demonstrating the similar trends for both the CG paths and longest wavelength in the p-band of the compound electronic spectra.

In the present study, we have re-examined the benzenoid CG's which were constructed proceeding from a number of topological rules on cyclicity [43]. By dealing with all isomer transformations upon which one benzene ring is transferred from one cata-fusion site to another, new CG's are obtained which incorporate considerably more directed edges. As an illustration, in fig. 9 the re-built CG(W) of all cata-fused benzenoids having 5 rings (fig. 8) is shown. It contains 30 edges corresponding to 30 isomer interconversions versus only 13 in the initial CG presented in ref. [1]. As can be seen in fig. 9, the CG reflects fairly well the trends in changing the longest wavelength λ in the p-band of the compound electronic spectra. Only in six cases is the respective molecular rearrangement associated with an increase instead of a decrease in λ: 5–8, 7–8 ($\Delta = +30$ cm^{-1}); 6–12, 11–12 ($\Delta = +50$); 9–11 ($\Delta = +85$); 9–12 ($\Delta = +135$). Thus, our comparability graph reproduces in 80% of cases the regularities in ordering these characteristic wavelengths. This may also be regarded as evidence for the importance of molecular topology for the spectral properties of benzenoid hydrocarbons. Other evidence for homologous benzenoids has been recently reported by Rouvray and El-Basil [63].

7. Some optimized structure-property correlations for C_9- and C_{10}-alkanes

In the two preceding sections, we have examined the degree to which various properties and characteristics of alkanes and benzenoids are conditioned by molecular topology or, more specifically, by molecular branching. Here, quantitative estimates are presented for the correlation that was supposed to exist between some alkane properties

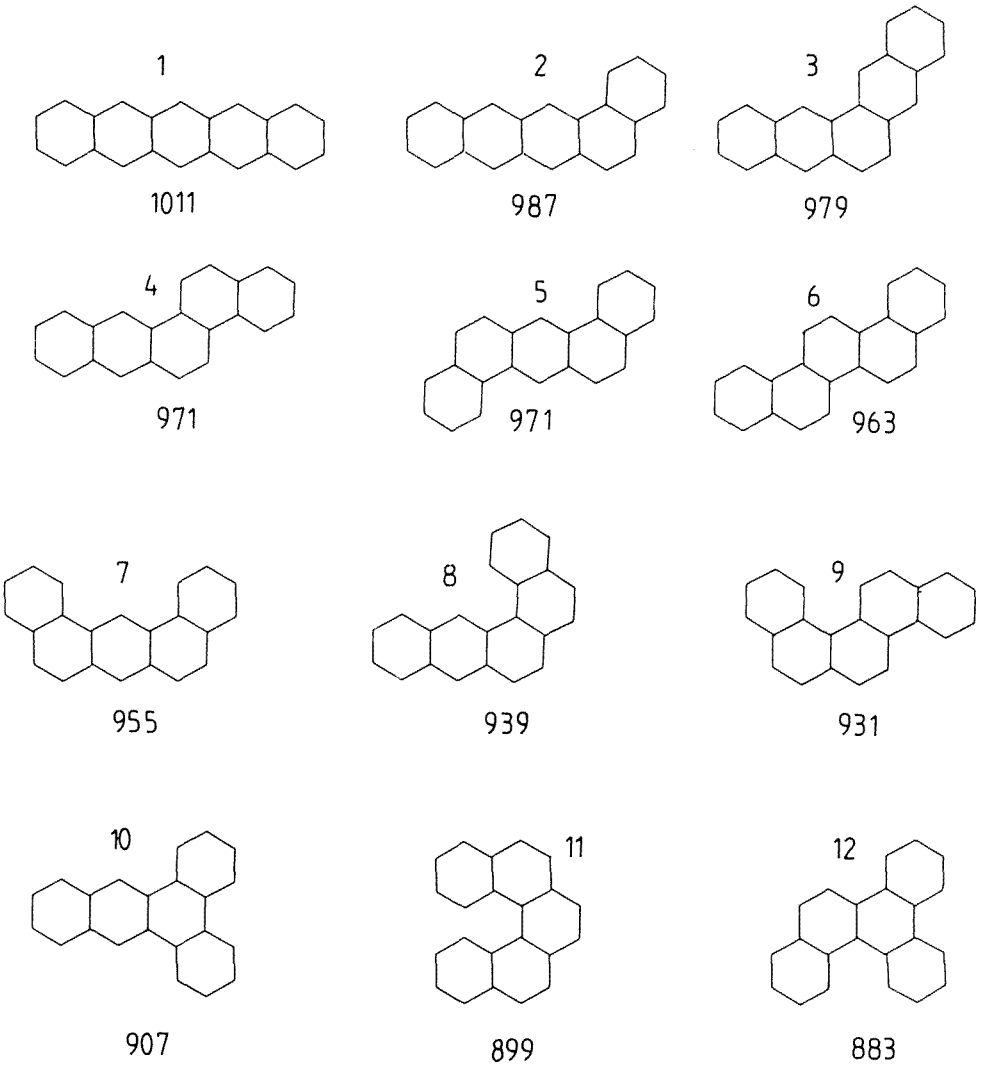


Fig. 8. Hydrogen-depleted molecular graphs of cata-fused benzenoid hydrocarbons with five rings and their Wiener number.

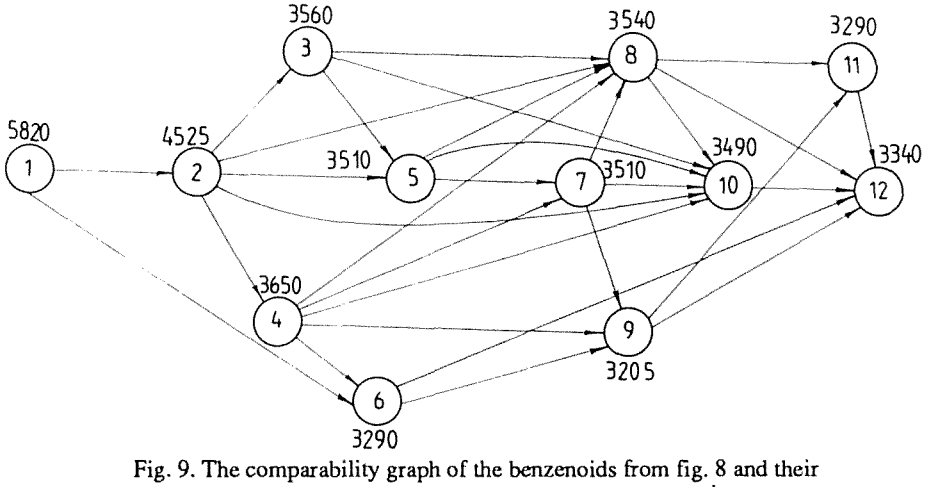


Fig. 9. The comparability graph of the benzenoids from fig. 8 and their largest wavelength in the electronic spectra p-band in cm^{-1} [62].

Table 3

Correlation between some thermodynamic properties [57] of C_9-alkanes and Wiener topological index

Correlation sample	Correlation equation	r	s	f
Critical densities[a]				
8, 13, 31, 35, 18	$0.3295 - 7.5123 \times 10^{-4}W$	0.9981	2.3×10^{-7}	3
1, 4, 5, 15, 35, 18	$0.3412 - 8.9481 \times 10^{-4}W$	0.9995	2.5×10^{-7}	4
1, 4, 21, 7, 17, 35, 18	$0.3432 - 9.1078 \times 10^{-4}W$	0.9997	1.0×10^{-7}	5
Critical volumes[b]				
8, 13, 31, 35, 18	$0.3568 + 1.4741 \times 10^{-3}W$	0.9979	9.9×10^{-7}	3
1, 4, 5, 15, 35, 18	$0.3247 + 1.8589 \times 10^{-3}W$	0.9994	1.1×10^{-6}	4
1, 4, 21, 7, 17, 35, 18	$0.3251 + 1.8531 \times 10^{-3}W$	0.9997	5.3×10^{-7}	5
Heats of vaporization[c]				
8, 13, 15, 35, 18	$29.8361 + 0.1417W$	0.9972	0.0123	3
1, 4, 5, 15, 35, 18	$31.7245 + 0.1188W$	0.9934	0.0530	4
1, 4, 21, 7, 17, 35, 18	$31.1095 + 0.1220W$	0.9710	0.2074	5
Entropy of vaporization[d]				
8, 12, 15, 35, 18	$71.7032 + 0.1378W$	0.9987	0.0056	3
1, 4, 5, 15, 35, 18	$73.3160 + 0.1187W$	0.9987	0.0104	4
1, 4, 21, 7, 32, 35, 18	$73.2239 + 0.1206W$	0.9860	0.0990	5
Critical pressure[e]				
8, 16, 30, 35, 18	$39.8642 - 0.1564W$	0.9936	0.0330	3
1, 2, 20, 3, 14, 34	$28.9527 - 0.0548W$	0.9135	0.1081	4
1, 4, 21, 7, 32, 35, 18	$36.8285 - 0.1241W$	0.9565	0.3391	5
Entropy in gas state[d]				
19, 16, 17, 35, 18	$183.0803 + 3.0718W$	0.9936	10.16	3
1, 2, 11, 26, 9, 33	$315.7597 + 1.6902W$	0.9920	8.98	4
Entropy in liquid state[d]				
19, 16, 17, 35, 18	$113.0539 + 2.6863W$	0.9937	7.64	3
1, 2, 11, 26, 9, 33	$235.0347 + 1.3127W$	0.9896	7.87	4

[a]In g/cm^3. [b]In l/mol. [c]In kJ/mol. [d]In J/mol.K. [e]In atm.

Table 4

Correlations between some thermodynamic properties [57] of C_{10}-alkanes and Wiener topological index

Correlation sample	Correlation equations	r	s	f
Critical densities[a]				
1, 2, 3, 25, 64	$0.3270 - 5.5446 \times 10^{-4}W$	0.9944	1.2×10^{-6}	3
1, 4, 7, 27, 66, 74	$0.3645 - 7.8327 \times 10^{-4}W$	0.9996	3.2×10^{-7}	4
1, 10, 12, 15, 67, 33, 74	$0.3620 - 7.6376 \times 10^{-4}W$	0.9991	5.0×10^{-7}	5
1. 10, 12, 15, 67, 65, 18, 75	$0.3619 - 7.6344 \times 10^{-4}W$	0.9973	1.3×10^{-6}	6
Critical volumes[b]				
1, 4, 38, 25, 62	$0.3599 + 1.4613 \times 10^{-3}W$	0.9929	1.1×10^{-5}	3
1, 4, 7, 27, 66, 74	$0.3243 + 1.6863 \times 10^{-3}W$	0.9995	1.6×10^{-6}	4
1, 2, 20, 49, 14, 34, 75	$0.3515 + 1.5636 \times 10^{-3}W$	0.9990	2.1×10^{-6}	5
1, 4, 38, 48, 67, 65, 18, 74	$0.3298 + 1.6519 \times 10^{-3}W$	0.9991	2.3×10^{-6}	6
Heats of vaporization[c]				
28, 30, 73, 71, 74	$27.7831 + 0.0777W$	0.9176	0.0491	3
8, 13, 54, 29, 72, 75	$24.9601 + 0.0919W$	0.9908	0.0309	4
1, 4, 12, 53, 32, 35, 74	$29.6702 + 0.0593W$	0.9942	0.0209	5
1, 10, 39, 15, 67, 65, 18, 74	$28.5603 + 0.0658W$	0.9833	0.0629	6
Entropies of vaporization[d]				
28, 30, 73, 71, 74	$58.2000 + 0.0200W$	0.9987	0.0053	3
55, 57, 67, 65, 18, 74	$60.3772 + 0.0201W$	0.9912	0.0349	4
1, 4, 7, 17, 65, 18, 74	$71.3661 + 0.1041W$	0.9884	0.1328	5
1, 36, 11, 26, 17, 65, 18, 74	$71.2550 + 0.1058W$	0.9830	0.1649	6
Critical pressure[e]				
1, 2, 3, 25, 64	$24.5250 - 0.0235W$	0.9700	0.0117	3
8, 13, 54, 29, 72, 75	$30.5550 - 0.0629W$	0.9849	0.0241	4
1, 2, 5, 15, 67, 33, 75	$30.2568 - 0.0592W$	0.9906	0.0348	5
Entropy in gas state[d]				
8, 42, 47, 9, 18, 74	$223.7265 + 2.1650W$	0.9959	9.773	4
1, 2, 37, 3, 9, 18, 74	$302.8512 + 1.4665W$	0.9903	24.038	5
Entropy in liquid state[d]				
8, 42, 47, 9, 18, 74	$135.8166 + 1.9869W$	0.9959	8.384	4
1, 2, 37, 3, 9, 18, 74	$213.6321 + 1.2947W$	0.9890	21.264	5

[a]In g/cm^3. [b]In l/mol. [c]In kJ/mol. [d]In J/mol.K. [e]In atm.

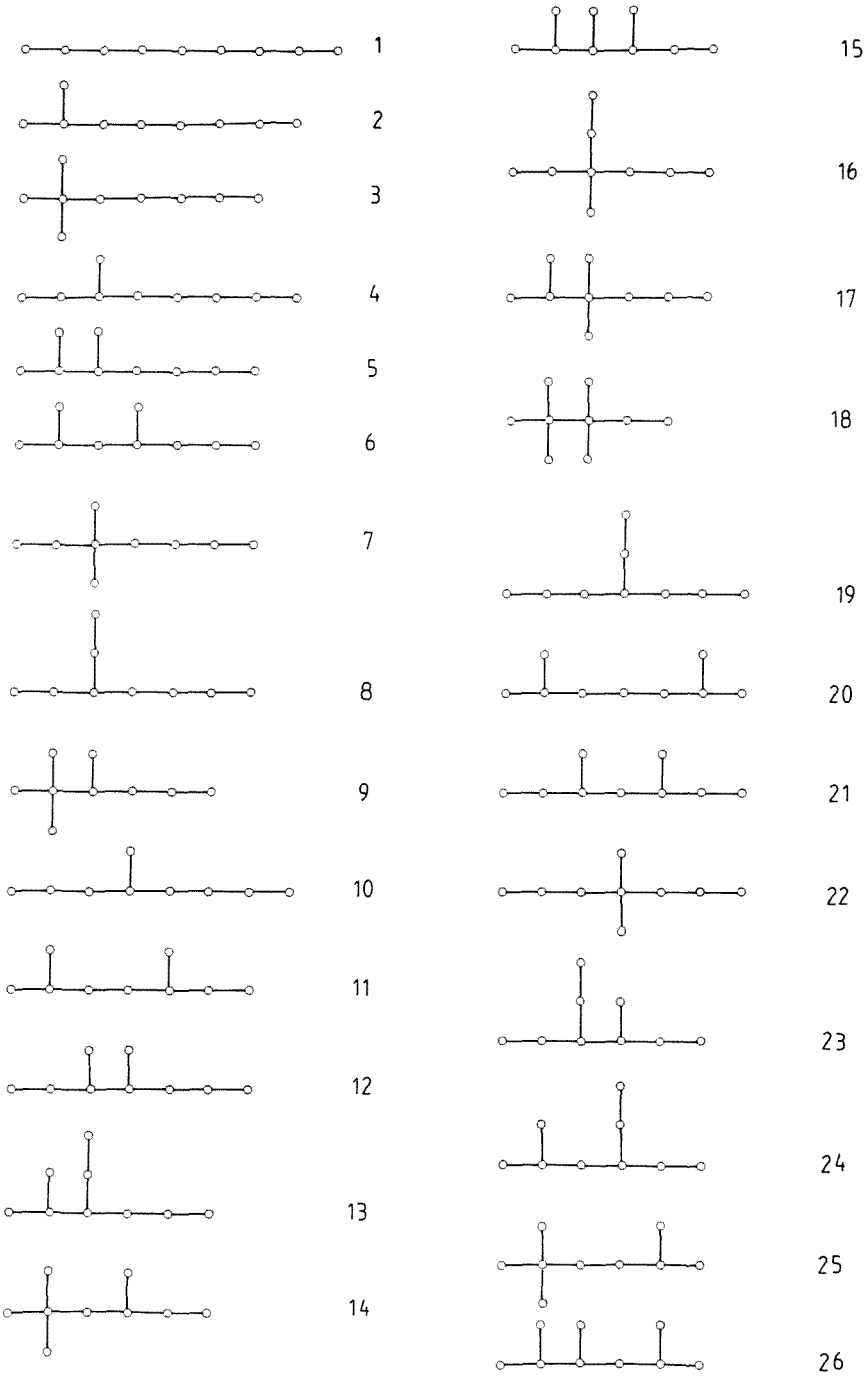


Fig.10.

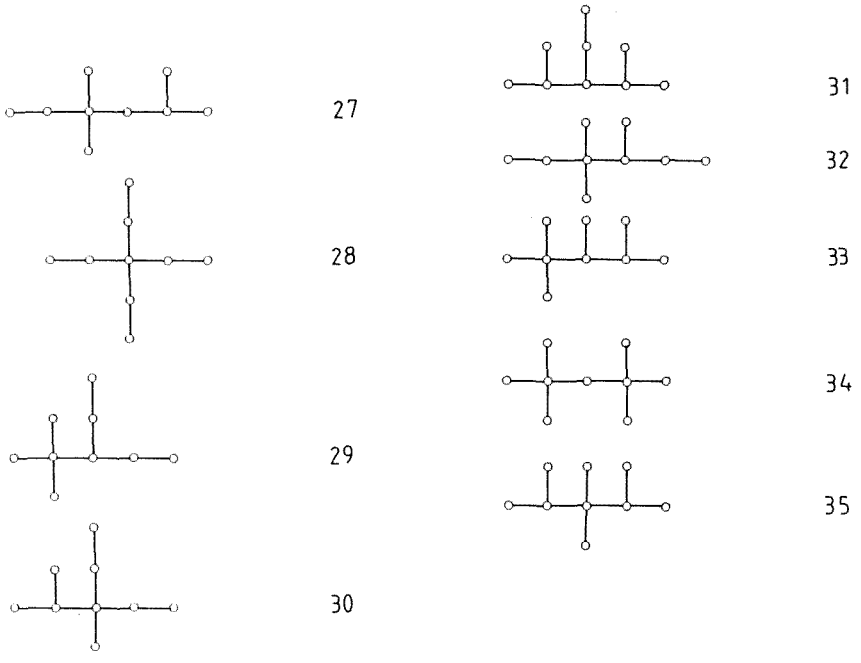


Fig. 10. Hydrogen-depleted molecular graphs of C_9-alkanes.

and the Wiener topological index utilized in constructing comparability graphs. This is a direct test for the main idea of the CG approach, according to which the regularities in molecular data are best expressed along the directed paths in the respective comparability graphs wherefrom the optimized correlation samples should be specified.

Seven C_9- and C_{10}-alkane properties were examined. Four of them, the heats and entropies of vaporization and entropies in gas and liquid state, belong to the group of alkane properties that follows very closely their regularities as produced by the comparability graphs. The other three properties, the critical pressures, volumes, and densities, belong to the second group of properties which also obey the CG regularities, but with a larger number of deviations as compared to the first group. The simplest linear correlation with the Wiener index was examined. The best equations thus obtained for CG paths of different length (5 to 8 isomers in each of them) are shown in tables 3 and 4, together with their correlation coefficients, standard deviations, and degrees of freedom. The isomer numbering is given in figs. 10 and 11, respectively.

As seen from these tables, the correlation obtained is very high, the standard deviation appearing as a more adequate estimate for the "quality" of correlation. Remarkably good linear correlations are obtained for critical densities and volumes, the standard deviations being down to the 10^{-5}–10^{-7} range. The next group of properties includes the heats and entropies of vaporization whose relative standard deviations are less than 0.15–0.25%. The critical pressures follow with $S_{rel} \leq 1.5\%$. The worst result

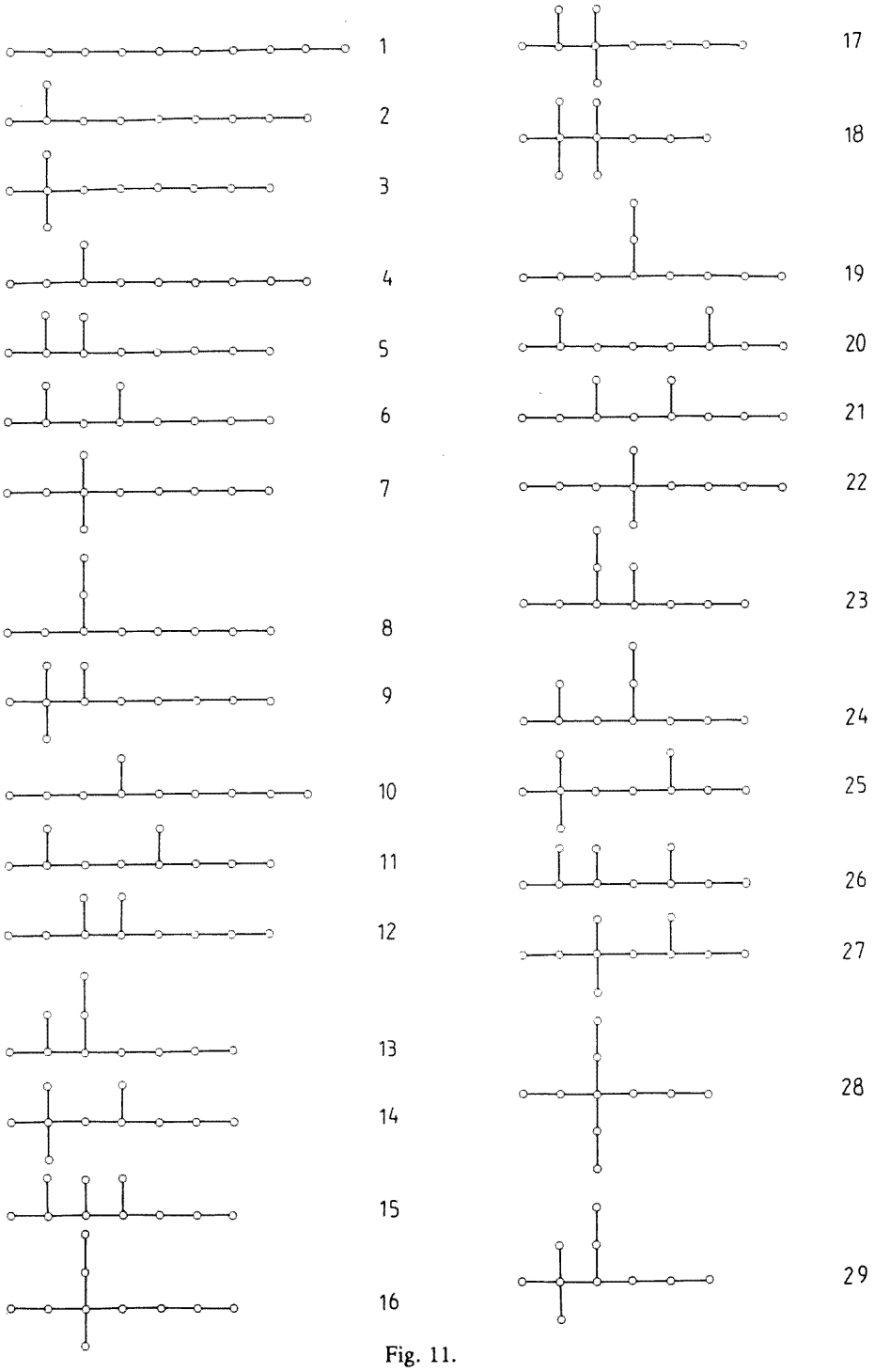


Fig. 11.

(See also following pages.)

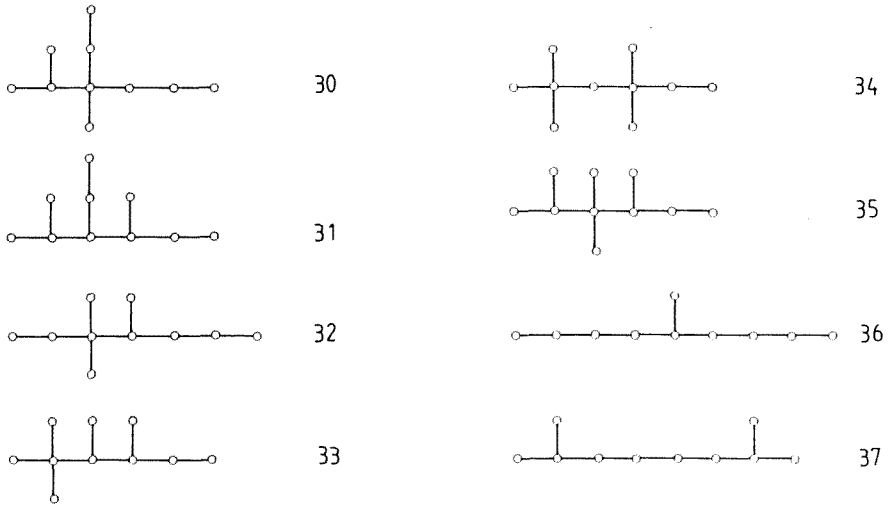


Fig. 11.
(See also following pages.)

is obtained for the two entropies: $S_{rel} \leq 3\%$ for C_9- and $S_{rel} \leq 7\%$ for C_{10}-alkanes. One may conclude that the Wiener topological index is a particularly suitable structural descriptor for calculating critical densities and volumes, as well as heats and entropies of vaporization. The C_9- and C_{10}-alkanes comparability graphs provide a sufficiently large number of optimized correlation samples for calculating these properties for all isomers. Thus, seventeen to nineteen equations were obtained for C_9-alkanes which have relative standard deviations of the four properties less than 0.001, 0.006, 0.6, and 0.3%, respectively. For the C_{10}-case, the number of equations is close to fifty and $S_{rel} \leq 0.005$, 0.02, 0.3 and 0.5%, respectively.

On the other hand, the unsatisfactory correlations found for critical pressures and entropies in gas and liquid state should not be attributed to the approach used. Rather, they indicate the need of nonlinear regression models, as well as the necessity of testing the other three topological indices used in the CG construction: the Randić and Hosoya indices, and the number of self-returning walks of degree 4. As an example demonstrating these possibilities for arriving at higher correlations, we have tested a randomly chosen sample of C_{10}-isomers (isomers 1, 4, 38, 7, 17, 65, 18, 74). The following equations resulted for their entropy in the liquid state:

$$S_1^0 = 208.0529 + 1.4029W, \qquad r = 0.949, \quad s = 96.3;$$
$$S_1^0 = 292.7688 + 1.6189Z, \qquad r = 0.953, \quad s = 89.6;$$
$$S_1^0 = -106.4940 + 110.2821\chi, \qquad r = 0.972, \quad s = 52.3;$$
$$S_1^0 = -2041.1663 + 962.2512\chi - 93.5395\chi^2, \qquad r = 0.986, \quad s = 33.0.$$

The improvement that can be achieved in this way is evident.

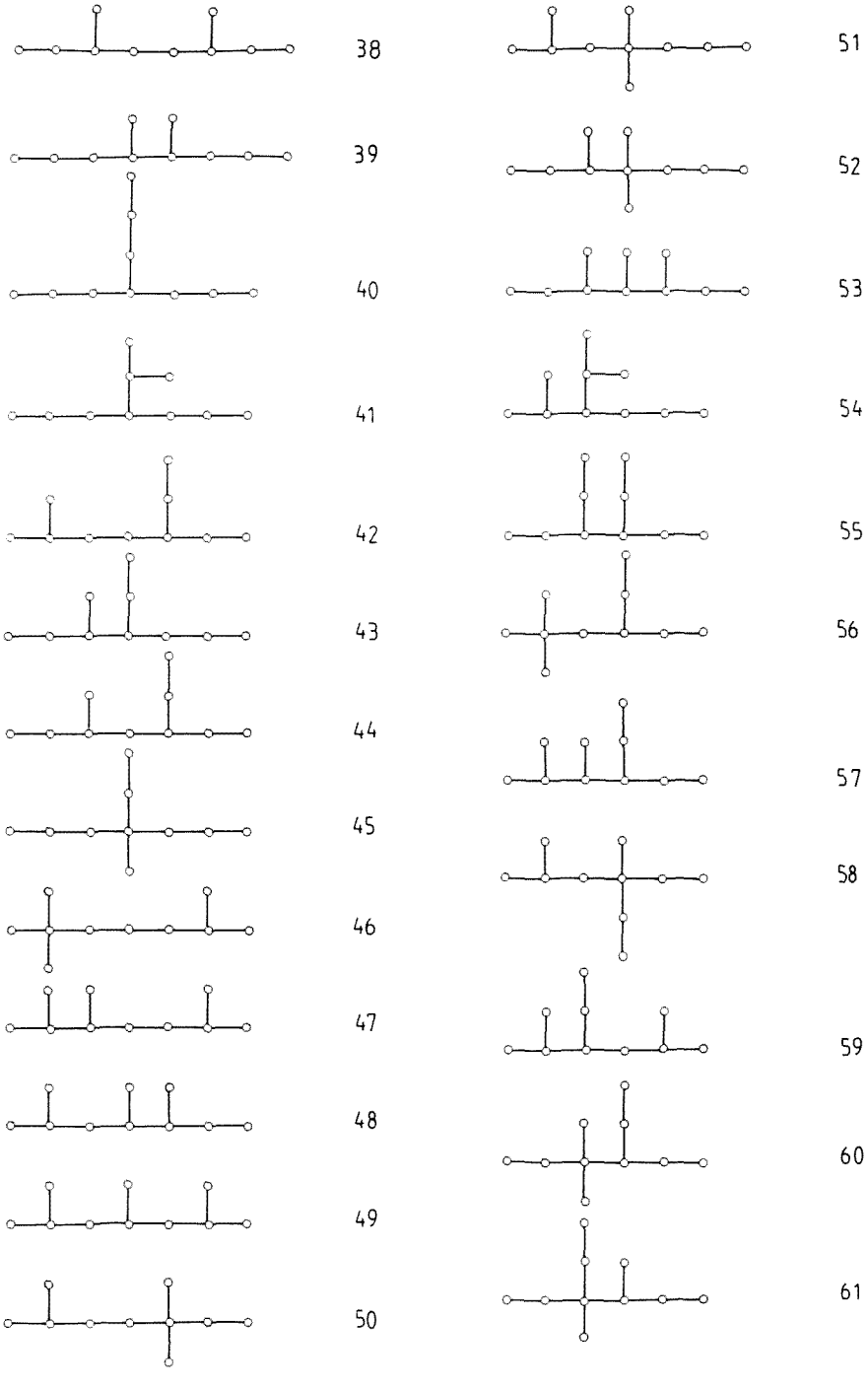


Fig. 11.
(See also following page.)

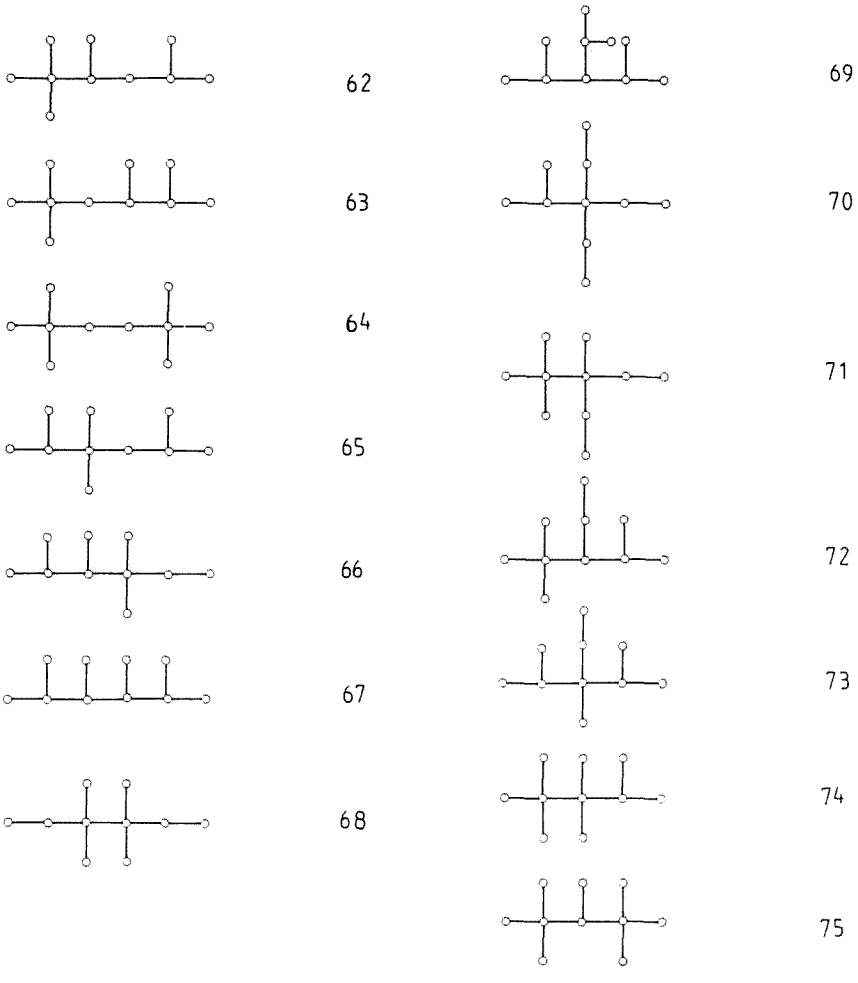


Fig. 11. Hydrogen-depleted molecular graphs of C_{10}-alkanes

8. Optimized correlation samples with variable number of atoms

The applicability of the comparability graph approach to the calculation of molecular properties depends crucially on the possibilities of predicting some unknown values of the properties. This is easily done by interpolation for a compound located in a CG where only part of the values are missing. In the general case, however, one needs to calculate the properties of higher homologues for which a few data are known. This raises the question of how to extrapolate the regularities found within a certain CG (i.e. for a constant number of atoms) beyond it. Otherwise, the question of a well-specified procedure of connecting different comparability graphs arises, so as to provide optimized CG paths with a variable number of atoms. Such a supercomparability graph (SCG) can be uniquely built up by the following *construction rule*:

An isomer belonging to a C_n-comparability graph can be connected with another compound included in the C_{n+1}-comparability graph if and only if the first compound can be converted into the second one by adding an atom.

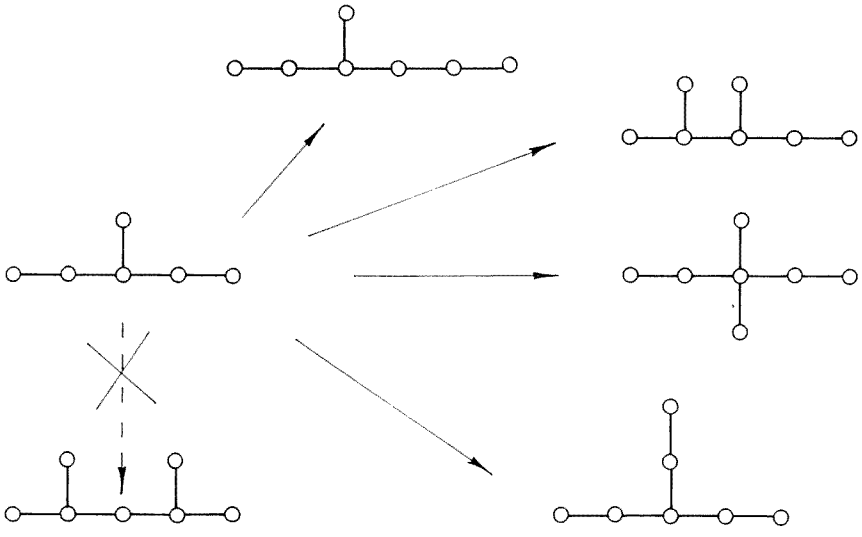


Fig. 12. Illustration of the construction rule for connecting the CG's: four allowed and one forbidden connections between C_6- and C_7-comparability graphs.

This is illustrated in fig. 12, where a C_6-alkane is shown to convert into four different C_7-isomers. Obviously, the fifth isomer cannot be obtained by our construction rule from the initial compound; hence, such a connection between C_6- and C_7-CGs is not possible.

In such a manner, we have built the supercomparability graph by connecting all alkane C_2 to C_{10} CGs. Its connection table is not given here for the sake of brevity. The number of directed paths in the SCG is very large indeed. This makes the search for the best correlations with a variable number of carbon atoms a time-consuming procedure which will be computerized along with the whole procedure in a subsequent study [55]. Here we present, in table 5, some selected examples of correlations between alkane properties (the critical densities, volumes, and pressures, heats of vaporization, and entropies in the gas state), the number of carbon atoms and the Wiener topological index. The number of compounds with a certain number of carbon atoms is denoted as a power series for each correlation, together with the compound total number.

Comparison of tables 5, 3, and 4 generally reveals the same trends for the correlations with a constant and variable number of carbon atoms. Again, the best results are obtained for critical densities and volumes, although the statistical estimates in the latter case are slightly worse. The alkane standard entropies are reproduced in the

Table 5
Thermodynamic properties [57] of C_2 to C_{10}-alkanes as a function of the number of carbon atoms and Wiener topological index

N	Correlation equations Number of compounds (total, as well as for each C_n)	r	s
1	$p_{cr} = 0.1060 + 0.0234n - 5.8036 \times 10^{-4}W$ $9(7^1 8^2 9^1 10^5)$	0.9817	6.1×10^{-6}
2	$p_{cr} = 0.0860 + 0.0274n - 7.4516 \times 10^{-4}W$ $10(7^1 8^2 9^1 10^6)$	0.9876	4.2×10^{-6}
3	$V_{cr} = 0.3319 - 2.2260 \times 10^{-3}n + 1.4437 \times 10^{-3}W$ $10(7^1 8^2 9^1 10^6)$	0.9974	1.3×10^{-5}
4	$V_{cr} = 0.0239 + 0.0576n - 1.7287 \times 10^{-3}W$ $13(2^1 3^1 4^2 5^1 6^1 7^1 8^1 9^4 10^1)$	0.9978	1.2×10^{-4}
5	p_{cr}[a] $= 22.7447 - 1.9634n - 0.1586W$	0.9456	0.628
6	$p_{cr} = 50.4103 - 4.9749n + 0.4082n^2 - 0.1465W$ $10(6^1 7^1 8^1 9^3 10^4)$	0.9943	0.079
7	$p_{cr} = 61.6573 - 8.0591n + 0.5587n^2 - 0.1102W$ $15(2^1 3^1 4^4 5^1 6^1 7^1 8^1 9^2 10^6)$	0.9975	0.406
8	$\Delta H_v = 151.1003 - 44.8711n - 2.6865n^2 + 0.0627W$ $10(7^1 8^2 9^1 10^6)$	0.9077	1.64
9	$\Delta H_v = 19.0221 + 13.6898n - 0.9397n^2 + 0.0101W$ $15(2^1 3^1 4^1 5^1 6^1 7^1 8^1 9^1 10^7)$	0.9946	1.53
10	$S_g^0 = 115.1588 + 65.0547n - 5.3482n^2 + 2.1367W$ $12(2^1 3^1 4^1 5^1 6^1 7^1 8^2 9^3 10^1)$	0.9990	9.10
11	$S_g^0 = 113.0480 + 66.8060n - 5.9013n^2 - 2.5384W$ $15(2^1 3^1 4^1 5^1 6^2 7^1 8^1 9^3 10^4)$	0.9990	9.45

[a]Equations (5) and (6) are derived for the same correlation sample.

same unsatisfactory manner. Critical pressures preserve their acceptable statistical estimates and even improve them when using a quadratic term for the number of carbon atoms. Worse are only the results for heats of vaporization. However, more definite conclusions will be possible only after the planned automated inspection of *all* possible paths in SCG [55].

9. Concluding remarks

The comparability graph concept is extended in this study along two important lines. The construction procedure is better specified, proceeding from the generalized Wiener number theory [56] which allowed us to deal with a considerably larger number of isomer interconvertions, including those associated with an intramolecular transfer

of more than one atom. The CGs thus built up incorporate many more paths, which are also generally longer than those from the CGs obtained previously [1–3]. Better chances are thus offered for specifying optimized correlation samples of isomeric compounds. From a theoretical viewpoint, the approach gains by the much more complete set of topological transformations taken into consideration, as well as by the representative set of graph invariants used in their quantitative characterization. The number of self-returning walks in the graph, which are related to the quantum-mechanical method of moments [49], is added in this study to the topological indices of Wiener, Randić, and Hosoya, representing the graph distance, connectivity, and non-adjacency, respectively. The optimized correlation samples specified by the paths in the combined comparability graph thus include isomers of which the systematic alteration in the moleclar structure is always associated with a regular change in the values of these four indices. Presumably, this could be a convenient basis for detecting regularities in molecular data. The examination made on twenty-seven alkane physico-chemical properties resulted in a division into three groups qualified as strongly dependent, dependent, and weakly dependent on molecular topology (or, more precisely, on the branching of the molecular skeleton). Thus, the regularities in sixteen alkane properties were found to be fairly well predicted by the comparability graphs. Aiming at the creation of a method for calculating molecular properties, a procedure is also specified for combining CGs with different numbers of vertices. Optimized correlation samples are thus obtained which include compounds with a variable number of atoms. The first structure-property correlations obtained on this basis revealed the Wiener number as the most convenient topological index for accurate calculations of the alkane critical densities and volumes, as well as entropies and heats of vaporization. All three results strongly prompt our further efforts toward the creation of a completely computerized system for CG contruction, their combination for variable numbers of atoms, and a search for the best structure-property correlations [55].

References

[1] D. Bonchev and O. Mekenyan, Chem. Phys. Lett. 98(1983)134.
[2] D. Bonchev and O. Mekenyan, J. Chem. Soc. Faraday Trans. 2, 80(1984)695.
[3] D. Bonchev, V. Kamenska and O. Mekenyan, Int. J. Quant. Chem., in press.
[4] P.G. Seybold, M. May and U.A. Bagal, J. Chem. Educ. 64(1987)575.
[5] L.B. Kier and L.H. Hall, *Molecular Connectivity in Chemistry and Drug Research* (Academic Press, New York, 1976).
[6] L.B. Kier and L.H. Hall, *Molecular Connectivity in Structure-Activity Analysis* (Wiley, New York, 1986).
[7] MATCH (Communications in Mathematical Chemistry), ed. O.E. Polansky (1975–1989).
[8] J. Math. Chem., ed. D.H. Rouvray (1987–1989).
[9] Discr. Appl. Math. 19(1988).
[10] *Annals of Mathematical Chemistry*, Vols. 1, 2, 3, ed. D. Bonchev and D.H. Rouvray (Gordon and Breach, London), in press.
[11] *Chemical Applications of Graph Theory*, ed. A.T. Balaban (Academic Press, London, 1976).
[12] N. Trinajstić, *Chemical Graph Theory*, Vols. 1, 2 (CRC Press, Boca Raton, FL, 1983).

[13] *Chemical Applications of Topology and Graph Theory*, ed. R.B. King (Elsevier, New York, 1983).
[14] A.T. Balaban, J. Mol. Struct. (Theochem) 120(1985)117; J. Chem. Inf. Comput. Sci. 25(1985)334.
[15] *Mathematics and Computational Concepts in Chemistry*, ed. N. Trinajstić (Horwood, Chichester, 1986).
[16] I. Gutman and O.E. Polansky, *Mathematical Concepts in Organic Chemistry* (Springer-Verlag, Berlin, 1986).
[17] *Graph Theory and Topology in Chemistry*, ed. R.B. King and D.H. Rouvray (Elsevier, Amsterdam, 1987).
[18] *Application of Graph Theory to Chemistry*, ed. G.S. Yablonskii (Nauka, Novosibirsk, 1988), in Russian.
[19] A. Sabljić and N. Trinajstić, Acta Pharm. Jug. 31(1981)183.
[20] A.T. Balaban, I. Motoc, D. Bonchev and O. Mekenyan, Topics Curr. Chem. 114(1983)21.
[21] D.H. Rouvray, Sci. Amer. 255(1986)40, Comput. Chem. 4(1987)470; Discr. Appl. Math. 19(1988)317.
[22] M.I. Stankevich, I.V. Stankevich and N.S. Zefirov, Usp. Khimii 57(1988)337.
[23] I.V. Stankevich, in ref. [18], p. 69.
[24] D. Bonchev, MATCH 7(1979)65.
[25] D. Bonchev, O. Mekenyan and N. Trinajstić, J. Comput. Chem. 2(1981)127.
[26] D. Bonchev and N. Trinajstić, Int. J. Quant. Chem. Symp. 16(1982)463.
[27] D. Bonchev, *Information-Theoretic Indices for Characterization of Chemical Structures* (Research Studies Press, Chichester, 1983).
[28] S.H. Bertz, in ref. [13], p. 206; Bull. Math. Biol. 45(1983)849.
[29] C. Raychaudhury, S.K. Ray, J.J. Ghosh, A.B. Roy and S.C. Basak, J. Comput. Chem. 5(1984)581.
[30] S.C. Basak, V.R. Magnuson, G.J. Niemi and R.R. Regal, Discr. Appl. Math. 19(1988)17.
[31] D. Bonchev, in: *Computational Chemical Graph Theory*, ed. D.H. Rouvray (Nova Publ., New York), in press.
[32] J.-E. Dubois, in ref. [11], p. 333.
[33] J.-E. Dubois and Y. Sobel, J. Chem. Inf. Comput. Sci. 25(1985)326.
[34] C. Mercier, I. Sobel and J.-E. Dubois, in: *Annals of Mathematical Chemistry*, Vol. 3, ed. D. Bonchev and D.H. Rouvray (Gordon and Breach, London), in press.
[35] J.W. Kennedy, M. Gordon and J.W. Essam, J. Chem. Soc. Faraday Trans. 2, 73(1977)1289.
[36] M. Randić, Chem. Phys. Lett. 55(1978)5215; Int. J. Quant. Chem. 5(1978)245.
[37] M. Barysh, N. Trinajstić and J.V. Knop, Int. J. Quant. Chem. Symp. 17(1983)441;
B. Jerman-Blazić, I. Fabić and M. Randić, J. Comput. Chem. 7(1986)176.
[38] M. Randić and C.L. Wilkins, J. Chem. Inf. Comput. Sci. 19(1979)31; Chem. Phys. Lett. 63(1979)1525.
[39] D. Bonchev and N. Trinajstić, J. Chem. Phys. 67(1977)4517.
[40] D. Bonchev, J.V. Knop and N. Trinajstić, MATCH 6(1979)21.
[41] D. Bonchev, O. Mekenyan, J.V. Knop and N. Trinajstić, Croat. Chem. Acta 52(1979)361.
[42] O. Mekenyan, D. Bonchev and N. Trinajstić, MATCH 6(1979)93.
[43] D. Bonchev, O. Mekenyan and N. Trinajstić, Int. J. Quant. Chem. 17(1980)845.
[44] O. Mekenyan, D. Bonchev and N. Trinajstić, Int. J. Quant. Chem. 19(1981)929.
[45] O. Mekenyan, D. Bonchev and N. Trinajstić, MATCH 11(1981)145.
[46] H. Wiener, J. Amer. Chem. Soc. 69(1947)17; J. Chem. Phys. 15(1947)766.
[47] S.H. Bertz, Discr. Appl. Math. 19(1988)65.
[48] D.H. Rouvray, in ref. [15], p. 295.
[49] J.K. Burdett, S. Lee and W.S. Sha, Croat. Chem. Acta 57(1984)1193;
J.K. Burdett and S. Lee, J. Amer. Chem. Soc. 107(1985)3050; in ref. [17], p. 302.
[50] M. Randić, J. Amer. Chem. Soc. 97(1975)6609.
[51] H. Hosoya, Bull. Chem. Soc. Japan 44(1971)2332.
[52] M. Randić, MATCH 7(1979)5.
[53] L. Lovasz and J. Pelikan, Period. Math. Hung. 3(1973)175.

[54] D. Bonchev, O. Mekenyan and O.E. Polansky, work in progress.
[55] D. Bonchev, O. Mekenyan and V. Kamenska, work in progress.
[56] O.E. Polansky and D. Bonchev, MATCH 21(1987)133.
[57] *Selected Values of Properties of Hydrocarbons and Related Compounds*, American Petroleum Institute Resrach Project 44 (Pittsburgh, Pennsylvania, 1977).
[58] R.C. Read, J.M. Prausnitz and T.K. Sherwood, *The Properties of Gases and Liquids* (McGraw-Hill, New York, 1977), Ch. 12.
[59] A.T. Balaban and I. Motoc, MATCH 5(1979)197.
[60] V.M. Tatevskii, *Fiziko-Khimicheskye Svoistva Individual'nikh Uglevodorodov* (Nauka, Moscow, 1960).
[61] A.T. Matukama, Proc. 7th Int. Symp. on Gas Chromatography, Copenhagen (1968), p. 55.
[62] E. Clar, *The Aromatic Sextet* (Wiley, New York, 1972).
[63] D.H. Rouvray and S. El-Basil, J. Mol. Struct. (Theochem) 165(1988)9.