

4

Combinatorial Enumeration in Chemistry

BY D. BABIĆ, D.J. KLEIN, J. VON KNOP AND N. TRINAJSTIĆ

1 Introduction

In this report we review the literature in the area of combinatorial enumeration in chemistry published in the last two years – from June 2001 to May 2003. In our earlier report¹ we presented historically important isomer enumerations, enumeration methods and then we reviewed the literature published between June 1999 and May 2001. Here we will focus only on the recent enumerative work since our previous report gives much background information on the history and methodology of counting objects in chemistry. Nevertheless, we will also mention some of the past achievements closely related to the discussed results most of which were not previously reviewed.

However, we wish to stress that the most fundamental problem underlying a part of modern chemistry called combinatorial chemistry^{2,3} is that a very large number of different molecules is possible. This fact raises the pertinent question *How many different molecules are possible?* It is a rather simple question, but it does not have a simple answer. Combinatorial enumerations can provide an answer.

2 Current Results

2.1 Isomer Enumeration. – In this section we will review the current results regarding enumerations of acyclic structures, usually represented by trees, and benzenoid hydrocarbons whose enumeration represents a subject considered by many past and present authors. We will also mention the related past results in order that the reader can see how the current achievement fits into the development of a particular area of isomer enumeration.

2.1.1 Trees. – The enumeration of all possible graph-theoretical trees⁴ and chemical trees⁵ still attracts attention. Trees are connected acyclic graphs.⁴ Chemical trees are trees in which no vertex has a degree greater than 4, whilst graph-theoretical trees do not have such a constraint. Chemical trees provide,

among many other things, the natural graph-theoretical representation of acyclic carbon structures. There has also been introduced by Kroto *et al.*⁶ in 1987 the concept of physical tree. Based on this concept a proposal was made for the simple mechanism by which acyclic molecules could be formed in interstellar space and circumstellar shells. Physical trees with no vertex-degree greater than 4 were generated and enumerated using the N -tuple code.^{7,8}

Physical trees carry ‘memory’ of their origin, implemented by a special labeling of their vertices, unlike chemical trees, that is, trees without ‘memory’. Thus, a physical tree is obtained by assigning labels to the vertices of a tree consecutively, and each vertex to be labeled must be adjacent to an already labeled vertex. Therefore, each vertex, except the vertex labeled 1, has exactly one neighbor with a lower label. This labeling results in the vertex-adjacency matrix⁵ A of a physical tree that contains only one non-zero element in each column of its upper (or in each row of lower) triangle.⁹ A consequence of the above is that the total number of different physical trees with N vertices and with no restriction upon the vertex-degree is simply $(N-1)!$. Since there are $(i-1)$ ways to pick the non-zero element in the i -th column of the upper (or in the i -th row of lower) triangle of the vertex-adjacency matrix, and since these choices are independent, the total number of ways equals $\prod_{i=2}^N (i-1) = (N-1)!$.¹⁰ The number of linear (chain-like) physical trees is also given by a simple formula:⁶ 2^{N-2} . For example, the number of physical trees with 13 vertices and the vertex-degree restriction to 4 or less is 311216626,⁸ whilst this number for all physical trees with 13 vertices is $12! = 479001600$. The number of linear physical trees with $N = 13$ is $2^{11} = 2048$. Physical trees represent a subset of labeled trees. The number of different labeled trees is N^{N-2} , a result given by Cayley already in 1889.¹¹ Hence, the number of all possible labeled trees with 13 vertices is $13^{11} = 1792159794037$.

Lukovits and Gutman¹² have shown that Morgan-trees represent a subclass of physical trees. The formal definition and nomenclature of a Morgan-tree was introduced by Lukovits¹³ in 1999 and has been named in honor of H.L. Morgan, who proposed the underlying concept in 1965, when working on the development of a computer-based chemical information system at the Chemical Abstracts Service in Columbus (Ohio), an algorithm for unique labeling of chemical structures.¹⁴

A Morgan-tree (MT) is a physical tree in which the numbering must obey an additional (with respect to physical trees) restriction: Each new label must be attached to a vertex which is adjacent to a vertex labeled with the lowest available ordinal. Example: Number 1 can be attached to any vertex. Number 2 must be attached to one of the vertices being adjacent to vertex 1, since an MT is also a physical tree. Number 3 must be attached to a vertex being adjacent to vertex 1, if there is such a vertex. If not, it must be attached to a vertex being adjacent to vertex 2, etc. The essence of the enumeration method based on MTs is that the number of MTs is a small fraction of physical trees. Any tree that can be labelled in such a way it is an MT. The codes representing MTs can be easily generated, and the redundant structures can be eliminated by using

several rules. Note that there is not any step which involves a comparison of different structures.

Lukovits and Gutman¹² derived a formula for enumerating all Morgan-trees (MT) with N vertices:

$$\text{MT}(N) = (2N-2)!/N! (N-1)! \quad (1)$$

They used Morgan-trees for the exhaustive generation and enumeration of chemical trees (alkane isomers and alkyl derivatives). In a previous work, Lukovits encountered a problem of redundant structures^{13,15} that appears now to be solved.^{12,16}

Došlić¹⁰ has shown that there is a simple bijection between the set of all Morgan-trees with a given number of vertices N and Dyck paths on $2(N-2)$ steps. A Dyck path on $2N$ steps is a lattice path in the coordinate plane (x,y) from $(0,0)$ to $(2N,0)$ with steps $(1,1)$ *up* and $(1,-1)$ *down*, never falling below the x -axis.¹⁷ Dyck paths are one of many combinatorial families enumerated by Catalan numbers (C_N) $N \geq 0$.¹⁸ These Dyck paths are also recognized as path diagrams (e.g. in ref. 19) corresponding to Yamanouchi symbols²⁰ which characterize spin-singlet coupling patterns for the symmetric group, as of use in many-particle quantum mechanics. A consequence of the said bijection is that the exact enumerative results for Morgan-trees with a given number of vertices can be obtained in terms of Catalan numbers. Došlić¹⁰ proved that the number of all Morgan-trees with N vertices is C_{N-2} . Lukovits-Gutman equation (1) is really equal to C_{N-1} . The seeming discrepancy between Lukovits-Gutman and Došlić formulas is due to the different starting points.

For a given $N \geq 0$, the quantity:^{18,21}

$$C_N = 1/(N+1) \binom{2N}{N} \quad (2)$$

is called the N -th Catalan number. Formula (2) generates Catalan numbers: 1, 1, 2, 5, 14, 42, 132, 429, 1430, 4862, 16796, 58786, 208012, etc. In Lukovits and Gutman's paper¹² these first thirteen Catalan numbers appear as matrix-elements next to the zero-diagonal elements of the (14×14) accessibility matrix S (see Table 1 in their paper) as the numbers of Morgan-trees with up to 13 vertices, but have not been identified by these authors as Catalan numbers.

The N -tuple code deserves a few more words. Most of the literature on isomer counting has been focused on the enumeration of specific structural classes. Very little attention has been given to the sequence of generation and its possible relationship to nomenclature. An exception to this trend is the N -tuple code developed for trees by von Knop *et al.*⁷ The N -tuple code represents a set of non-negative integers smaller than N (the number of vertices in a tree; the number of carbon atoms in an alkane), each representing the degree of a vertex in a tree or subtree. The degree (or valency) of a vertex in a (molecular) graph is equal to the number of edges meeting at this vertex. To

obtain the N -tuple code, one has first to identify the vertices with the highest degree and select amongst them one that will result in a code that produces lexicographically the largest number. After the starting vertex is located, that vertex and adjacent edges are removed. The subtrees thus produced are examined. Typically, this means looking for the largest chain, and, if several chains of the same length appear, their codes are derived and combined in such a way that the result corresponds to the lexicographically highest number. The N -tuple code was the basis for the so-called *compact* codes²² (called so because they use a limited number of digits for linearly encoding a given chemical structure) that have the potential to develop into a full new system of chemical coding. Compact codes have been used to encode^{23–27} various classes of saturated and unsaturated acyclic molecules including acyclic compounds with heteroatoms, polycyclic molecules such as benzenoid hydrocarbons, annulenes, aza-annulenes, annulenoannulenes, aza-annulenoannulenes, cyclazines, aza-cyclazines, etc. In Figure 1 we give as an illustrative example the N -tuple code for a tree representing the carbon skeleton of 2,2,3-trimethylpentane and the labeling of its vertices induced by the N -tuple code.

The N -tuple code was also extended by Contreras *et al.*²⁸ to polycyclic structures and other molecular structures of increasing complexity. This code together with new approaches for symmetry analysis and cycle detection in molecular graphs was implemented in a modular way in the program CAMGEC2.²⁸

Into this area of enumeration there also fall the efforts of Davidson^{29–31} who used side-chain complexity minimization algorithms for naming and coding carbon skeletons of alkanes and ring-chain assemblies. His latest effort³² is focused on fast canonical generation of the alkane series. Davidson's approach is based on the selection of the main chain of an alkane as the path that yields the least complex side chains without the maximum-length constraint that

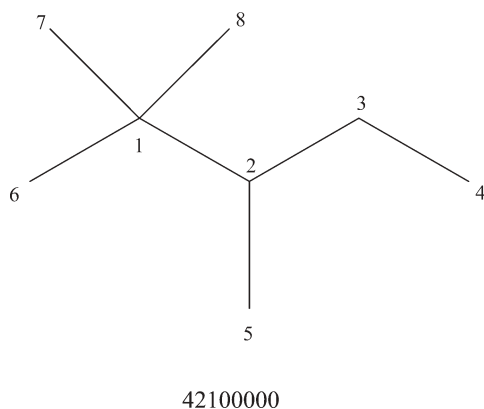


Figure 1 A tree representing the carbon skeleton of 2,2,3-trimethylpentane, its N -tuple code and the labels of its vertices induced by the N -tuple code

leads to an efficient algorithm representable as a nested binary tree. The largest side-chain required to specify an N -carbon alkane becomes $(N-1)/3$. This, for example, allows 3.8 million C_1 - C_{22} alkanes to be coded for name translation in dictionary order, using an alphabet of 33 C_1 - C_6 alkyl groups, also ranked by complexity. The generating process produces reversible isomer codes, making the computation rate in isomers per second inverse linear with N and much faster than reported rates for other structure-generators. Davidson called his alkane isomer codes (size + alkyl codes) the *ultimate* compact codes. We note that the number of isomers for C_{22} alkanes, given by Davidson (2267998),³² is smaller than given by other authors (2278658).^{8,33}

Classical enumerations of the constitutional isomers of various classes of hydrocarbons in the manner of Henze and Blair³⁴⁻⁴⁰ are still being pursued. Henze and Blair derived recursive algorithms for enumerating the number of constitutional isomers of acyclic structures using the corresponding radicals. For example, they computed isomers of alkanes using alkyl radicals. However, Henze and Blair did not tackle the enumeration of constitutional isomers of cyclic structures. Lam from the Department of Chemical Pathology of the Chinese University of Hong Kong (China) undertook this task since he realized that alkyl polyradicals can be employed to enumerate isomers of cyclic compounds using ideas, and even numbers, of Henze and Blair. After showing that alkyl 1,1-biradicals can be used to enumerate the constitutional isomers of alkenes⁴¹ (these numbers fully agree with the numbers obtained by von Knop *et al.*^{7,8} using an algorithm based on the N -tuple code) and alkyl-cyclopropanes,⁴² Lam⁴³ used alkyl 1,1,1-triradicals to enumerate isomers of alkylcyclobutadienes. For an alkyl triradical of carbon content N , the number of constitutional isomers is L_N , where also $L_N = S_{N-1}$ where S_{N-1} denotes the number of constitutional isomers of an alkyl group with $N-1$ carbons. The S_N numbers Lam took from Henze and Blair.³⁴

Wang *et al.*⁴⁴ have dealt with enumeration of hydroxyl ethers. The considered class of compounds was limited to the so called *stable* hydroxyl ethers, with at most single oxygen atom bonded to any carbon atom. They presented generating functions for constitutional isomers, all stereo isomers and achiral stereo isomers of these compounds.

2.1.2 Benzenoid Hydrocarbons. – The interest in answering the ubiquitous question *How many benzenoid hydrocarbons are there?* continues. In the last forty years, this question was considered by many authors who proposed a galaxy of methods for generation and enumeration of benzenoid isomers^{8,45,46} (in the mathematical literature this problem is known as counting hexagonal polyominoes^{47,48} or hexagonal animals⁴⁹ and belongs to the cell-growth problems^{4,50}). Graphs representing benzenoid hydrocarbons are often called in the literature polyhexes,^{5,51} benzenoid graphs^{5,52} and there is also a term polyhex hydrocarbons^{53,54} in use. Polyhexes are planar graphs that may be obtained by any combination of regular hexagons such that any two of its hexagons have exactly one common edge or are disjoint. These schemes sometimes differ in

whether they count polyhex structures which are not subgraphs of the honeycomb lattice (i.e., they sometimes do and sometimes do not include helicenic structures).

Among the first attempts in the chemical literature to enumerate benzenoid structures were efforts by Balaban and Harary⁵¹ in 1968, Balaban *et al.*⁵⁵ in 1980 and von Knop *et al.*⁵⁶ in 1983. With increasing capacity of computers and an improvement of algorithms, enumerations of benzenoids have been extended to very large systems. After von Knop and his group published results for simply- and multiply-connected benzenoids with up to 16 hexagons,^{53,57} using an algorithm based on the DAST (dualist angle-restricted spanning tree) code, several groups extended these enumerations, using their own algorithms, gradually building up to systems with up to 25 hexagons, *e.g.* refs. 58–63. All these efforts belong to the so-called *constructive* enumerations, that is, counting benzenoids by constructing each molecule using a suitable code. The *constructive* enumerations are useful for generating combinatorial libraries.

In order to extend the enumerative results to much higher numbers of hexagons, Vöge *et al.*⁶⁴ devised a new algorithm, based on methods used in statistical physics, which allows the computer enumeration of benzenoid systems (without constructing them; the so-called *nonconstructive* enumeration) with up to 35 hexagons. They proved that $b_h \approx \kappa^h$, where b_h is the number of benzenoid systems with h hexagons and κ is a growth constant. They also established the rigorous bounds for the growth constant: $4.789 \leq \kappa \leq 5.905$ and estimated that $\kappa = 5.16193016(8)$.

The work of Vöge *et al.*⁶⁴ made it possible to establish the asymptotic law by which b_h increases with h , for large values of h .⁶⁵ The numbers b_h for $h > 10$ turn out to be rather huge, *e.g.*, $b_{11} = 141229$, $b_{16} = 359863778$, $b_{21} = 1012565172403$, and $b_{35} = 5851000265625801806530$.⁶⁴ Using the asymptotic form one can calculate the approximate values of b_h for h greater than 35, if one is interested in these numbers. However, the exact computations were rather lengthy – Vöge *et al.*⁶⁴ needed about 10 weeks on a Compaq Alpha-Server ES40, utilizing up to 5 GB of memory, to get the values of b_h with up to $h \leq 35$. Even so, this is considerably less time than one used by von Knop *et al.*⁵⁷ to get the values of b_h with up to $h \leq 16$ – they needed almost 92 days on a personal computer Siemens PCD3D (20 MHz, 386-AT).

It should be also mentioned that the number of known benzenoid hydrocarbons is about a thousand. However, the methods for synthesis of benzenoids have been improved^{66,67} so that even very large benzenoids⁶⁸ are already available exhibiting unusually high stability^{68,69} and many interesting properties (*e.g.*, self-organization, conductivity).⁷⁰ There is also available a strategy, based on graph-theoretical and topological arguments, for designing all kinds of aromatic structures that includes benzenoid hydrocarbons and their derivatives.⁷¹

Klein and Misra⁷² addressed a problem of enumerating poly-*N*-phenacenes, consisting of a cycle of benzene rings each fused in either *meta* or *para* fashions to the rings on either side of it. Here the fusion may be at opposite edges of a ring (termed *para*-fusion) or at next-neighbor edges of a ring (and termed

meta-fusion). Thence a cycle of rings results, where the *meta*-fusions can be to non-neighbor edges either to the “left” or “right”, as one proceeds around the cycle. Note here may be made of synthetic efforts toward preparing and characterizing such species.^{73–77} The enumeration of the possible isomers is reminiscent of the Pólya approach, which entails enumeration of suitable symmetry classes of mathematical maps from a domain (usually of skeletal sites) to a range (of substituents). However for the present enumeration it turns out the relevant symmetry group acts both on the domain and range, so that some extension of the standard enumerative approach is needed. Thence a general symmetry-attentive method for enumeration is made, and applied for the cases where the number N of benzene rings is prime. The method is more general than prime N , this being the simplest case, where the relevant cyclic-related groups have simpler subgroup structures. The methodology is first presented on the simpler case of open-chain polyphenacenes, as earlier treated by Balaban and Harary.⁵¹ Leading up to the general approach a systematic scheme to generate the various isomer labels is formulated, is noted to relate to a systematic ‘nomenclature’, and applied to cycles of up to 13 benzene rings (where there are over a 30000 isomers, of the ‘regular’ type). The method and applications incorporate the possibility of the cyclo-phenacene being ‘regular’ like an (untwisted) belt, or with a Möbius twist (either left- or right-handed), or even being multiply twisted. Some cursory attention is even directed to the possibility of knotted cyclo-polyphenacenes (possibly with various numbers of twists) and polyphenacenes that consist of a linking of multiple cycles.

2.1.3 Pólya Theory and Mark Tables. – The classification of substitutional isomers according to their symmetry realized as a result of the substitution has been a topic of some chemical interest for a little over 2 decades, with several chemical researchers utilizing methodology parts of which dates back to Burnside⁷⁸ around 1900. A number of researchers (Hässelbarth,⁷⁹ Mead,⁸⁰ Brocas, Gielen and Willem,⁸¹ and Fujita⁸²) utilized so-called *mark tables* to solve this type of problem, such tables conveniently characterizing the different possible subgroup structures within the parent permutationally represented symmetry of the skeleton into which the different substituents are being placed. Over the years books on this subject have appeared, by Fujita⁸³ and more recently by El-Basil,⁸⁴ merging the whole idea neatly with conventional Pólya theory of enumeration.⁸⁵ In 2002, a special issue of MATCH (vol. 46) was published dedicated to recent developments in this area. Thence the bulk of these papers may all be nicely reviewed together here. The introductory paper there by El-Basil⁸⁶ gives a short (dozen-page) history, with emphasis on the important role played by Pólya theory and mark tables in counting isomeric structures.

Dolhaine and Hönig^{87,88} continue their work on counting isomers of complex organic compounds using the *Mathematica*[®] application program package *Isomers*,⁸⁹ which is the implementation of an efficient isomer enumeration algorithm based on the classical Pólya counting theorem.⁸⁵ They applied their

program to obtain the number of possible tetramers of a selected subset of the nine known inositols (1,2,3,4,5,6-hexahydroxycyclohexanes). The question *How many tetramers of four specific inositols, namely neo-, muco-, D-chiro and L-chiro are there?* has been asked by Tomas Hudlicky from the University of Florida at Gainesville, whose group is making these compounds.^{90,91} This question was directed to Dolhaine and his coworkers because Dolhaine *et al.*⁹² applied in 1999 their program *Isomers* to a number of molecules of interest for organic chemists, including some oligomers of inositols. The answer to this question should also be of interest in the field of structural biology, where sugars and pseudosugars gain status as information-carrying macromolecules.⁹³ Doing combinatorial chemistry without knowledge of just how many combinations are to be expected, appears to be impractical, or at least incomplete. Dolhaine and Hönig⁸⁷ obtained these numbers – there are 24300 *linear* and 6480 *branched* tetramers. Therefore, there are 30780 tetramers possible which are made of the four specific and different (*neo-*, *muco-*, *D-chiro* and *L-chiro*) inositols.

In a subsequent publication, Dolhaine and Hönig⁸⁸ gave tables of all possible combinations and stereoisomers of the known nine inositols in their dimer, trimer and tetramer forms. The respective numbers of achiral compounds and cumulative estimations of the number of possible linear pentamers, hexamers, heptamers, octamers and nonamers are also listed. These calculations have been carried out by the *Mathematica – AddOn ‘Isomers.m’* program which is available together with some help, free of charge (for academics only), at: <http://www.cis.TUGraz.at/orgc/institut/softnew.htm>.

Kerber⁹⁴ presents a mathematically elegant concise overview of the fundamentals of enumeration theory under finite group action (following up on his mathematics book⁹⁵). Kerber attends but briefly to the realized and potential chemical applications, so that his overview may be seen as encompassing an indication of further techniques of likely chemical use. In the following article Hässelbarth and Kerber⁹⁶ describe applications of enumeration under finite group action on a particular class of combinatorial libraries that arise from a symmetric parent compound by means of reactions with a given set of building blocks. Such libraries were described by Carell *et al.*,^{97,98} and their enumeration by weight was made by van Almsick *et al.*,⁹⁹ using the Pólya theorem.⁸⁵ The enumeration by symmetry group is discussed by Hässelbarth¹⁰⁰ in 1986. In their paper Hässelbarth and Kerber⁹⁶ give a refinement of both these methods, enumerating such libraries by weight and by symmetry group using mathematical apparatus presented by Kerber in the preceding paper.⁹⁴ They present their methodology on previously considered libraries,^{97,98} arising from xanthene and from a benzene skeleton with three acid-chloride groups.

Nemba and Balaban¹⁰¹ obtained the number of chiral and achiral isomers of cycloalkanes C_NH_{2N} with M homomorphic alkyl groups. The M alkyl groups C_NH_{2N+1} are homomorphic if they are structurally and stereochemically identical when they are detached from the ring system. Nemba and Balaban¹⁰¹ used the upgraded form of the enumeration technique described by Robinson *et*

al.,¹⁰² and produced numbers of enantiomeric pairs and achiral carbon skeletons of homopolysubstituted cycloalkanes $C_N H_{2N-M} (C_N H_{2N+1})_M$ for various values of $N \geq 3$ and $1 \leq M \leq N$.

In the mentioned special issue of MATCH, there are general methodological papers by Fujita and El-Basil¹⁰³ and by Lloyd.¹⁰⁴ A paper by Klein and Misra⁷² is mentioned earlier in section 2.1.2 here. Another paper by Lloyd and Dolhaine¹⁰⁵ applies the mark-table methods outside of the area of isomer enumeration – enumerating *thermal cycles*.

Fujita⁸³ continues an immense near single-handed work on the development of symmetry-attentive combinatorial isomer-enumeration methodology and its application. In the present work, Fujita¹⁰⁶ reports exhaustive combinatorial enumeration of the square-planar complexes with achiral and chiral ligands under the point-group D_{4h} , where its mark table^{86,103} and its inverse mark table are prepared and used to calculate the subduction of coset representations. The enumeration results are used to discuss equivalency under point-group symmetry, that is, enantiomeric relationships for chiral complexes and prochirality for achiral complexes. The alternative combinatorial enumeration of the square-planar complexes was carried out under the permutation group S_4 , which is the symmetric group on four indices.¹⁰⁷ After defining the proper and improper permutations, the subgroups of S_4 are classified into stereogenic and astereogenic groups for the square-planar complexes. Then, equivalency under permutation-group symmetry is employed to clarify enantiomeric and diastereoisomeric relationships. Fujita's work based on the comparison between the action of a point-group and that of a permutation-group provides a novel and versatile approach to stereochemistry.

Fujita¹⁰⁸ also studied molecules of ligancy 4 that have been derived from allene, ethylene, tetrahedral, and square-planar skeletons to show that their symmetries are dually and distinctly controlled by point groups and permutation groups. He has shown the following: inasmuch as the point-group symmetry controls the chirality/achirality of a molecule, sphericity in a molecule¹⁰⁹ and enantiomeric relationship between molecules,¹⁰⁹ the permutation-group symmetry controls the stereogenicity of a molecule, tropicity in a molecule and diastereoisomeric relationship between molecules. To characterize permutation groups, proper and improper permutations have been defined by comparing proper and improper rotations. These permutation groups were classified into stereogenic and astereogenic groups. After a coset representation (CR) of a permutation group has been ascribed to an orbit (equivalence class), the tropicity of the orbit has been defined in terms of the global stereogenicity and the local stereogenicity of the CR. As a result, the conventional stereogenicity has been replaced by the concept of local stereogenicity in this paper. Fujita coined the terms homotropic, enantiotropic, and hemitropic and has used them to characterize prostereogenicity. A molecule is defined as being prostereogenic if it has at least one enantiotropic orbit. Since this definition has been found to be parallel with the definition of prochirality, relevant concepts have been discussed with respect to the parallelism between stereogenicity and chirality in order to restructure the theoretical foundation of stereochemistry and stereoisomerism. The derivation of the skeletons has

been characterized by desymmetrization due to the subduction of CRs. The Cahn-Ingold-Prelog (CIP) system¹¹⁰ has been discussed from the permutational point of view to show that it specifies diastereomeric relationships only. The apparent specification of enantiomeric relationships by the CIP system has been shown to arise from the fact that diastereomeric relationships and enantiomeric ones overlap occasionally in case of tetrahedral molecules.

Fujita¹¹¹ has extended the sphericity concept, proposed for specifying stereochemistry in a molecule,¹⁰⁹ to a study of stereoisomerism among molecules. The novelty of this approach is to characterize the global symmetries of molecules as the *local symmetries of stereoisomerism*. Therefore, stereochemistry and stereoisomerism have been discussed on a common basis. Promolecules, which have been generated as stereochemical models of molecules by placing proligands (structureless ligands with chirality/achirality) on the vertices of a tetrahedral skeleton, have been analyzed by a permutation-group approach and by a point-group one. The skeleton has been considered to belong to the symmetric group of degree 4 and to the isomorphic point group T_d . The chirality fittingness derived from the sphericity concept has been applied to the characterization of local symmetries of promolecules, where two types of the Young tableaux have been compared. Thus, the Young tableaux of symmetry have been introduced to treat the ligand packing based on the chirality fittingness. These tableaux have been compared with the Young tableaux of permutation, which have taken no account of such chirality fittingness. The two types of the Young tableaux have been applied to the enumeration of tetrahedral isomers with chirality fittingness observance and without chirality fittingness. This enumeration has enabled the clarification of the quantitative aspect of the sphericity concept in characterizing isomer equivalence. Therefore, equivalent isomers under a point-group symmetry have been shown to construct an orbit of stereoisomers that is ascribed to a coset representation. Homomeric, enantiomeric, and diastereomeric relationships between stereoisomers have been discussed by means of homospheric, enantiospheric and hemispheric orbits of stereoisomers. Skeleton-based and ligand-based categories for enantiomers and diastereomers have also been discussed. The stereogenicity and the prostereogenicity of the Chan-Ingold-Prelog system¹¹⁰ have been related to the Young tableaux of permutation.

Fujita and El-Basil¹¹² have developed a graphical method for generating one- and some two-dimensional group characters by using a reduced homomer set. The reduced homomer set was obtained by introducing the notion of negative graphs as corresponding to homomers with exchanged colors of vertices.

2.2 Kekulé Structure Count. – As remarked in our previous report¹ chemical work on counting Kekulé structures appears to be less wide-spread than a decade or two ago. It is however still studied by the mathematicians^{113,114} and sometimes by chemists.¹¹⁵ In graph-theoretical literature Kekulé structures are identified as 1-factors,¹¹⁶ perfect matchings^{52,113,117} or Kekulé patterns.^{114,118,119} In 1985 von Knop *et al.*⁴⁵ published a book including a comprehensive listing

of benzenoids of up to 9 hexagons along with their Kekulé-structure counts. Cyvin and Gutman¹¹⁹ published a book in 1988 on methods for counting Kekulé structures in benzenoid hydrocarbons. Dias¹²⁰ a year earlier (1987) produced a handbook on benzenoid hydrocarbons that contained for each structure also its Kekulé number.

Misra and Klein¹²¹ have further considered theoretical characteristics of cyclo-polyphenacenes, including the enumeration of their Kekulé structures. Such species always have Kekulé structures, and a comparable enumeration algorithm was made a while back by Cyvin *et al.*,¹²² though they focused more on structures like kekulene¹²³ – that is, this earlier work focused on *unstrained* structures. The enumeration proceeds by way of a transfer-matrix technique,^{5,124–128} such as discussed in our previous report.¹ This technique has most usually previously been applied to polymers with but a single monomer unit, whereas in the present case there are three monomer units (*para*-fused, left-*meta*-fused, and right-*meta*-fused) occurring in different sequences. As occasionally previously noted¹²⁷ the technique applies reasonably to such *disordered* polymers also. Misra and Klein¹²¹ incorporate the Kekulé-structure counts into a reactivity index, which in application to open-chain polyphenacenes is found to agree well with an available set of experimental data.¹²⁹ Further they utilize a simple *combinatorial curvature* index to indicate the amount of strain in the structures, to obtain decent agreement with Dobrowolski's earlier SCF results¹³⁰ on all 52 of the regular cyclo-hexaphenacenes.

Torrens¹¹⁵ considered the enumeration of Kekulé structures for alternant benzenoid hydrocarbons using the connection between the permanent of the vertex-adjacency matrix and the number of Kekulé structures:

$$\text{per } A = K^2 \quad (3)$$

This relation holds only for alternant structures as early noted by Percus.¹³¹ It was discussed in graph-theoretical terms and applied to alternant hydrocarbons by Cvetković *et al.*^{132,133} In the above formula, K is the number of Kekulé structures and $\text{per } A$ is the permanent of the vertex-adjacency matrix A of an alternant hydrocarbon (*i.e.* a bipartite graph,⁴ or bicolourable structure¹³⁴).

Permanents are generally more difficult to compute than determinants,¹³⁵ so that Torrens¹¹⁵ proposed an algorithm for fast computation of the permanents of sparse matrices. Note, the vertex-adjacency matrix ([0,1] matrix) is for molecular graphs usually a sparse matrix. Torrens' approach based on an earlier algorithm reported by Cash,¹³⁶ applies the algorithm in an upgraded form to a suitable submatrix of the vertex-adjacency matrix. If we color an alternant structure with N sites by two colors in such a way that sites colored with the same color are never connected, then the sites can be divided in two groups each containing sites of the same color. The corresponding vertex-adjacency matrix is consequently given in terms of two submatrices B and B^T , each submatrix connecting between sites of different colors:

$$A = \begin{bmatrix} \mathbf{0} & \mathbf{B} \\ \mathbf{B}^T & \mathbf{0} \end{bmatrix} \quad (4)$$

where \mathbf{B}^T is the transpose of \mathbf{B} and $\mathbf{0}$ is the zero matrix. Then, the permanent of the adjacency matrix is given by:

$$\text{per } A = \text{per } \mathbf{B} \text{ per } \mathbf{B}^T = [\text{per } \mathbf{B}]^2 \quad (5)$$

From (3) and (5) follows:

$$K = \text{per } \mathbf{B} \quad (6)$$

In fact all these equations have been known for several years, *e.g.* refs. 132–134. Torrens' contribution then is a computer program based on (6) which is about seven times faster than the C-language program devised by Cash,¹³⁶ part of Torrens' program's advantage coming from the utilization of bitwise operations instead of integer comparisons. Still among the molecules considered by Torrens are two structures (pentalene, azulene) that are non-alternant hydrocarbons, though they have an alternant perimeter.

Lukovits *et al.*¹³⁷ have calculated the number of Kekulé structures of a special class of carbon nanotubes (CNT) with open ends which are also called tubulenes. CNTs are composed of cylindrical graphene sheets consisting of sp^2 carbon atoms and may be thought of as hexagonal conjugated systems surrounded by a cloud of π -electrons. They were first prepared by Iijima¹³⁸ in 1991 as tiny tubes with about 1.5 nm of diameter and a length of several microns. A tube of this diameter contains about 20 hexagons along its circumference. CNTs possess conducting or semiconducting properties depending on the tube structure^{139–142} and may be prepared by a laser technique¹³⁸ or by electrochemical procedures¹⁴³ or in some other way, *e.g.* refs. 144–148.

Lukovits *et al.*¹³⁷ have calculated the number of Kekulé structures in defect-free *armchair* tubulenes¹⁴⁹ using the transfer-matrix technique.^{1,5,124–128} Other researchers have also been engaged in generating the number Kekulé structures of nanotubes and tubulenes, *e.g.* refs. 1, 150–153.

Lukovits *et al.*¹³⁷ considered polynaphthalenes and polyphenanthrenes and their macrocyclic analogs. The results they obtained indicate that the number of Kekulé structures is greater if the considered tubulenes are extended in the vertical (*i.e.*, if they get longer) than in the horizontal direction. Moreover, experimental observation obtained by using scanning probe microscopy supports this prediction.¹³⁷ But here it should be emphasized that this conclusion is for the considered tubulenes, with the particular types of ends considered by Lukovits *et al.*¹³⁷ That is, the global Kekulé structure count can be vastly different depending on the ends of the tubes, because of a *long-range* order inherent in Kekulé structures. That the Kekulé-structure count is very sensitive to the ends is evident with the trivial example entailing the addition (or subtraction) of but a single vertex from the end of a Kekulé-structure-rich polymer, whence with an odd number of sites the count drops abruptly to 0. But a more

general understanding is available.¹⁵⁴ The resultant general form of the count $K_{L,c}$ for a polymer of length L and circumference c appears asymptotically (for large L) as

$$K_{L,c} \approx A_c A_c^L \quad (7)$$

where A_c and A_c are independent of L . Further A_c (the maximum eigenvalue of the transfer matrix) for a given type of tube ends up dependent on the ends of the polymer tube only in a rather simple discretized fashion. That is, of the infinitude of possible ends there are but $\approx c$ different classes, each class giving a distinct value of A_c . Overall for general conjugated π -network polymers (non-zero) A_c falls in a range between 1 and σ^c , with σ independent of c and L as well as the particular monomer unit. For a particular polymer this full range from $1 \rightarrow \sigma^c$ is typically approximately uniformly filled (by $\approx c$ values). In any event the simple relation of Swinborne-Sheldrake *et al.*¹⁵⁵ correlating overall resonance energy with the Kekulé structure count that is employed by Lukovits *et al.*¹³⁷ should be used with caution for extended systems if one uses a global (*i.e.*, end-sensitive) Kekulé structure count.

The sensitivity in selected properties of general graphitic fragments to the nature of different boundaries, has been studied sometime back,¹⁵⁶⁻¹⁵⁸ as this relates to the tendency for unpaired electrons to appear in extended graphitic nano-structures. That is, with an unfavorable boundary which gives very few Kekulé structures, it becomes more favorable for major contributions from resonance structures with unpaired electrons at the boundary if the unpaired electrons are so placed that the remnant effective boundary is such as to support a greater number of Kekulé structures. Indeed it has been found (earlier on in Klein¹⁵⁹ and Klein and Bytautas,¹⁶⁰ and recently now by Ivanciuc *et al.*^{161,162}) that the predictions made from such relatively simple resonance-theoretic considerations end up correlating quite accurately with the number of unpaired boundary-localized electrons predicted by suitable UHF solutions to Hubbard models. In these articles such comparisons have been made for about two-dozen different graphitic edges, much extending several earlier pieces of previously observed but unrationalized behaviors for the Hückel-model solutions at boundaries of graphite¹⁶³⁻¹⁶⁵ or of large benzenoids.^{157,158,166} Further in the work of Ivanciuc *et al.*^{161,162} results have been enunciated for a general class of vacancy defects, and also in this case there has been found agreement in about two dozen cases with UHF solutions for the Hubbard model in regard to the occurrence of defect localized unpaired spins. Especially for the case of a single-site vacancy defect the results here again rationalize some earlier observations¹⁶⁷⁻¹⁷⁰ made with simple tight-binding methods. Further there is a qualitative agreement with the results of scanning electron microscopy for two types of edges,¹⁷¹ and for the single-site vacancy defect.¹⁷²⁻¹⁷⁴

2.3 Counting Walks. – Walks can be used for characterization¹⁷⁵⁻¹⁷⁹ of (molecular) graphs, for quantification¹⁸⁰ of their complexity (or more precisely their labyrinthicity¹⁸¹) and for definition of various molecular descriptors.^{177,182-184} A

walk in a (molecular) graph is an alternating sequence of vertices and edges, such that each edge begins and ends with the vertices immediately preceding and following it.⁴ The length of the walk is the number of edges in it. Repetition of vertices and/or edges is allowed in a walk. A random walk in a graph is naturally designated by a suitable probability measure, of which there are two natural ones. First, one probability measure entails starting walks from each vertex with equal likelihood and subsequent steps are such that each neighboring vertex is stepped to with equal probability, so that the probability of stepping from vertex i to j is $1/d_i$ where d_i is the degree of vertex i . Second, another conceivable probability measure takes each possible walk of a given length as equally likely. If a graph is regular (*i.e.*, having all vertices of the same degree), then the two resultant probability distributions for walks are the same – but not otherwise. The walks generated with the first probability measure are typically called *random*, and here we term the walks of the second case *equi-weighted*. A self-returning walk is a (random) walk starting and ending at the same vertex.

It is well-known^{5,185,186} that the number of walks of length ℓ beginning at vertex i and ending at vertex j is given by the i,j -element of the ℓ -th power of the vertex-adjacency matrix A : $(A^\ell)_{ij}$. Then the number of self-returning walks of length ℓ is given by the i,i -element of the ℓ -th power of the vertex-adjacency matrix A : $(A^\ell)_{ii}$. It has been found, for example, thirty years ago, that the total number of self-returning walks of length ℓ which coincide with the trace of A^ℓ can be applied to the theory of total π -electron energy,¹⁸⁷ in as much as this gives moments of the eigenvalue distribution (as elaborated some later here). And more recently, self-returning-walk counts have been found to be applicable to systematic search of isocodal vertices in molecular graphs.^{175,188–191} Isocodal vertices in a graph are those vertices that have the same numbers of self-returning walks for each length of walk.

The atomic walk count of order ℓ , denoted by $(awc)_\ell$, is the number of all possible walks of length ℓ which start at a specified vertex (atom) i and end at any vertex (atom) j :¹⁸⁰

$$(awc)_\ell(i) = \sum_{j=1}^N (A^\ell)_{ij} \quad (8)$$

where N is the number of vertices in a graph G .

The molecular walk count of order ℓ , denoted by $(mwc)_\ell$, is obtained by summing up all atomic walk counts of order ℓ :^{182,192}

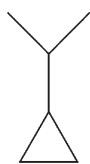
$$(mwc)_\ell = \sum_{i=1}^N (awc)_\ell(i) = \sum_{i=1}^N \sum_{j=1}^N (A^\ell)_{ij} \quad (9)$$

The total walk count, denoted by twc , is the sum of all $(mwc)_\ell$ for $\ell = 1$ to $N - 1$:^{182,192}

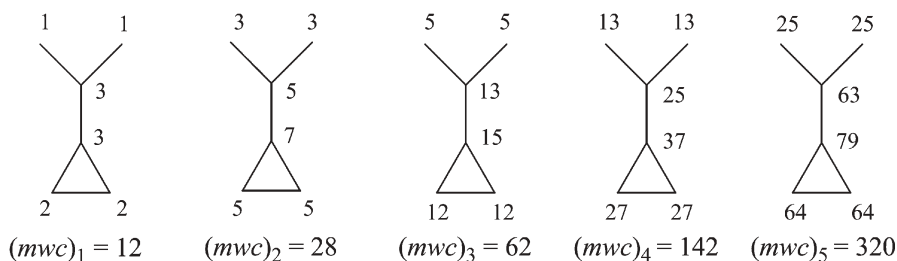
$$twc = \sum_{\ell=1}^{N-1} (mwc)_{\ell} = \sum_{\ell=1}^{N-1} \sum_{i=1}^N \sum_{j=1}^N (A^{\ell})_{ij} \quad (10)$$

The total walk count can be obtained in a much simpler way utilizing a relationship between the Morgan extended connectivities¹⁴ (EC) and matrix powers of the vertex-adjacency matrix. This relationship was first observed by Razinger,¹⁹³ who gave an empirical example but offered no proof. Rucker and Rucker¹⁸⁰ proved that EC 's and $(awc)_{\ell}$'s are identical. Consequently, the EC of vertex i of the ℓ -th order can be obtained by iterative summation of the $(\ell - 1)$ -th order contributions of the neighbors of i and is equal to $(awc)_{\ell}(i)$. Thus, calculation of twc requires nothing but a sequence of addition steps. This elegant procedure, termed by Rucker and Rucker¹⁸⁰ the Morgan summation procedure, is illustrated in Figure 2. Rucker and Rucker¹⁸⁰ also prepared

(a) Graph representing the carbon skeleton of isopropylcyclopropane



(b) Values of $(awc)_{\ell}(i)$ obtained by iterative summation over all neighbours, starting from vertex-degrees and a values of $(mwc)_{\ell}$



(c) The value of twc

$$twc = 12 + 28 + 62 + 142 + 320 = 564$$

Figure 2 An illustrative example of the Morgan summation procedure

program MORGAN (so named presumably to point to the pioneering work of Morgan¹⁴) based on the described procedure.

There is also a third method available for counting total walks. This method is based on the relationship between molecular walks of length ℓ , eigenvectors (c_{ri} ; $i=1,2,\dots,N$) and eigenvalues λ_r^ℓ of the vertex-adjacency matrices with various values of exponent ℓ .¹⁹² A starting point is the eigenvalue equation:

$$A\mathbf{c}_r = \lambda_r\mathbf{c}_r; r = 1, 2, \dots, N \quad (11)$$

Using the standard matrix-theoretical methods, from eq. (11) follows:

$$(A^\ell)_{ij} = \sum_{r=1}^N \lambda_r^\ell c_{ri} c_{rj} \quad (12)$$

Eq. (12) in combination with eq. (9) gives:

$$(mwc)_\ell = \sum_{r=1}^N \sigma_r \lambda_r^\ell \quad (13)$$

where

$$\sigma_r = (c_{r1} + c_{r2} + \dots + c_{rN})^2 \quad (14)$$

If λ_r is a degenerate eigenvalue, then σ_r , as defined by (14), is not uniquely determined. If such a case occurs, then the sum of σ_r 's over all degenerate eigenvectors is a true graph invariant σ'_ρ with ρ a label for distinct eigenvalues. Then in place of (13) one has $(mwc)_\ell = \sum_\rho \sigma'_\rho \lambda_\rho^\ell$. Formula (13) represents the spectral decomposition of the molecular-walk counts.^{116,194,195}

Rücker and Rücker¹⁸³ reported total walk counts for saturated acyclics through polycyclic hydrocarbons with up to ten carbon atoms. They also reported walk counts for molecules containing multiple bonds and/or heteroatoms using the Morgan summation procedure.^{181,196} Graphs which are used to represent this kind of molecule are *general* graphs,¹⁹⁷ that is, graphs in which multiple edges and/or loops are allowed. In *simple* graphs (as we earlier considered) multiple edges and loops are *not* allowed.¹⁹⁸ General graphs are also known as the vertex- and edge-weighted graphs.⁵ Walks in general (weighted) graphs depend on weights given to vertices and edges.^{181,196,199} For example, if a loop on vertex i is given a weight of 1, it means that it will give rise to additional walks such as the walk of length 1 from i to i . Once we decide the values for vertex- and edge-weights, the application of the Morgan summation procedure is straightforward. In Figure 3 we give the total walk count for a selection of molecules taken from Rücker and Rücker.¹⁹⁶

In fact the bulk of the discussion concerning the counting of equi-weighted walks carries over with but slight modification to the generation of random

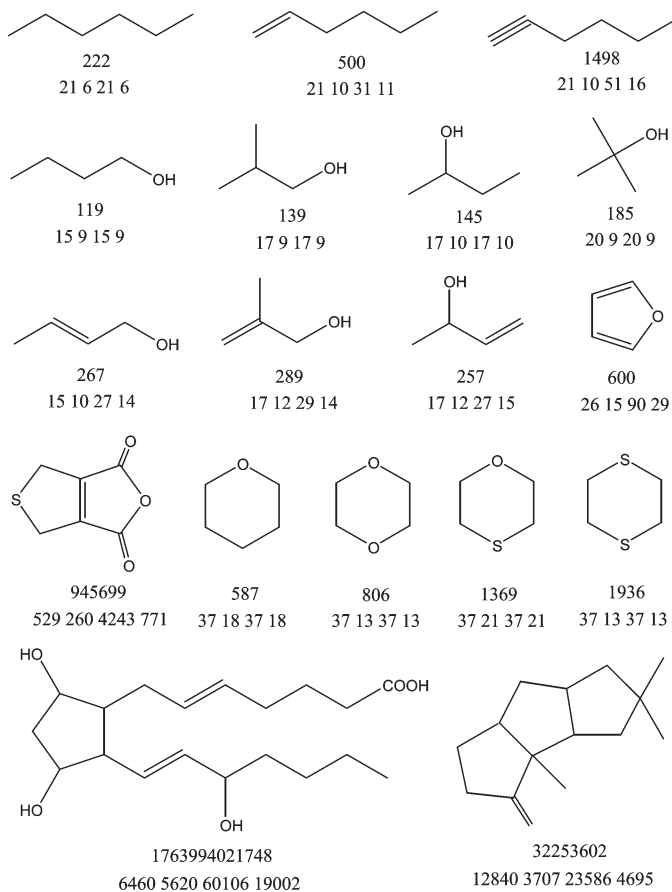


Figure 3 Total walk counts (*twc*) for a selection of molecules. *O* and *S* are arbitrarily assigned weights 1 and 2, respectively. Four numbers in the row below *twc* are the total number of connected substructures (N_c), the number of distinct connected substructures (N_s), the total number of connected subgraphs (N'_c) and the number of distinct connected subgraphs (N'_s). Data presented here are all taken from ref. 196

walk sums. Random walks are generated by powers of a (Markov) matrix M with elements which are probabilities for the associated steps

$$(M)_{ji} = \begin{cases} 1/d_i, & j \neq i \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

Then e.g. $(M^\ell)_{ji}$ is the probability for an ℓ -step random walk beginning at site i to end at site j . In analogy to the $(awc)_\ell(i)$ invariant one may consider a random-walk count:

$$rwc_{\ell}(i) = \sum_{j=1}^N (M^{\ell})_{ij} \quad (16)$$

which has the neat interpretation that $rwc_{\ell}(i)/N$ is the probability that one ends after ℓ steps at site i having started randomly at any site. The sum $\sum_{i=1}^N rwc_{\ell}(i)$, which would be the analog of $(mwc)_t$, is not so interesting though in that this is always $= N$, because this sum divided by N is evidently the probability of ending someplace after ℓ steps. (And thence also the analog of twc is trivial also.) Again the $rwc_{\ell}(i)$ may be neatly computed via eigensolutions to the matrix M , which however generally is non-symmetric, when the graph is not regular. But in this general case the eigensolutions are related to those of a symmetric matrix, as may be seen if one first starts out noting that M may be written as $M = AD^{-1}$ where D is the diagonal matrix of degrees d_i . Then we introduce the matrix:

$$H = D^{-1/2} M D^{1/2} \quad (17)$$

where $D^{1/2}$ is the diagonal matrix of square-roots of degrees and $D^{-1/2}$ is its inverse. Evidently H is similar to M (*i.e.*, they are related by a similarity transformation), so that H has the same eigenvalues, and eigenvectors \mathbf{c}'_r which are simply related to those \mathbf{c}_r of M , thusly

$$\mathbf{c}'_r = D^{-1/2} \mathbf{c}_r \text{ and } \mathbf{c}_r = D^{1/2} \mathbf{c}'_r \quad (18)$$

But evidently also $H = D^{-1/2} A D^{-1/2}$, so that H is symmetric and more standard matrix diagonalization routines may be employed, to yield the \mathbf{c}_r and associated eigenvalues λ_r . Then *e.g.*,

$$(rwc)_{\ell}(i) = \sum_r (c_{ri} d_i^{-1/2}) \lambda_r^{\ell} \left(\sum_{j=1}^N c_{rj} d_j^{-1/2} \right) \quad (19)$$

where the c_{ri} are components of \mathbf{c}_r . In fact this matrix H is rather well-known, sometimes being termed the (*normalized*) *Laplacian* matrix of G , and there is much theory about it, *e.g.*, as reviewed in Chung.²⁰⁰ This matrix is also interesting in that half the sum of its elements is just the Randić²⁰¹ connectivity index ${}^1\chi$. In general one may anticipate that many different graph invariants defined in terms of equi-weighted graphs may also have interesting and useful analogues defined in terms of random graphs. Some such invariants have been noted in by Klein²⁰² to be related to *sum rules* for resistance distances. But there are many graph invariants built for equi-weighted graphs where the random-walk analogues have not yet been explored. One example of this is the different invariants derived by Diudea^{203,204} from his *walk matrices*.

Dress *et al.*²⁰⁵ reported the number of walks in trees with up to 18 vertices. The total number of trees they considered was 205004 trees. They studied the

harmonicity²⁰⁶ of these trees and found that exactly three of the studied trees are harmonic, none almost harmonic, 11 superharmonic, and all other considered trees to be subharmonic. Dress and Gutman²⁰⁷ also reported asymptotic results regarding the number of walks in a graph.

The concept of walks was extended by Lukovits and Trinajstić²⁰⁸ to walks of zero and negative orders by using a backward algorithm based on the usual procedure to obtain the values of molecular walk counts. These authors called the Morgan summation procedure the Razing algorithm, and have formalized it as follows. Let $\mathbf{1}$ be an N -dimensional (column) vector, all entries of which are equal to one. Let \mathbf{a}_ℓ be a (column) vector, the entries of which are the awc 's of order ℓ for all vertices. Then the Razing algorithm can be formally written as:²⁰⁹

$$\mathbf{a}_\ell = \mathbf{A}^\ell \mathbf{1} = \mathbf{A}^{\ell-1} \mathbf{a}_1 = \dots = \mathbf{A} \mathbf{a}_{\ell-1} \quad (20)$$

It is clear if matrix \mathbf{A}^{-1} exists, the procedure can be inverted:

$$\mathbf{a}_{\ell-1} = \mathbf{A}^{-1} \mathbf{a}_\ell \quad (21)$$

Formula (21) allows the concept of atomic walk counts to be extended to awc 's of zero and negative orders. However, there are three possible subclasses of the inverse procedure: (I) The inverse of the vertex-adjacency matrix exists – this is the simplest case; (II) matrix \mathbf{A}^{-1} does not exist, and $\mathbf{a}_0 = \mathbf{1}$ and (III) matrix \mathbf{A}^{-1} does not exist, and $\mathbf{a}_0 \neq \mathbf{1}$. Lukovits and Trinajstić²⁰⁸ have shown how to get walks of negative order for cases (II) and (III). Note that awc 's and mwc 's of negative order may assume non-integer and even negative values. In Figure 4 we give $(awc)_\ell$ and $(mwc)_\ell$ for several values of ℓ starting with $\ell = 1$ up to $\ell = -5$ for ethylcyclobutane.

Rücker and Rücker,²¹⁰ after learning about the work by Lukovits and Trinajstić,²⁰⁸ derived a nonrecursive equation for walks of zero and negative orders in a graph (or molecule), utilizing their previous work on spectral decomposition.^{192,195} They also developed a computer program for counting the atomic and molecular walks for both positive and negative orders. With these two papers^{208,210} the concept of walk counts covers walking in both directions – forward and backward, and no walking at all.

2.4 Combinatorial Measures of Molecular Complexity. – Most authors agree that the concept of complexity is too complex to be defined precisely.^{211–213} And it has been further proposed that complexity rather than being a non-numerical quantity is a partially ordered quantity – see, footnote 27 of the ref.²¹⁴ But for a chemist's purpose one may consider a system to be complex if it consists of a great number of components – the more components in a molecule, the more complex the molecule is.

We do not fully review here the variety of existing complexity indices and their dependencies on structural features (size, branching, cyclicity, multiple edges, heteroatoms, symmetry) of molecular graphs (they were recently

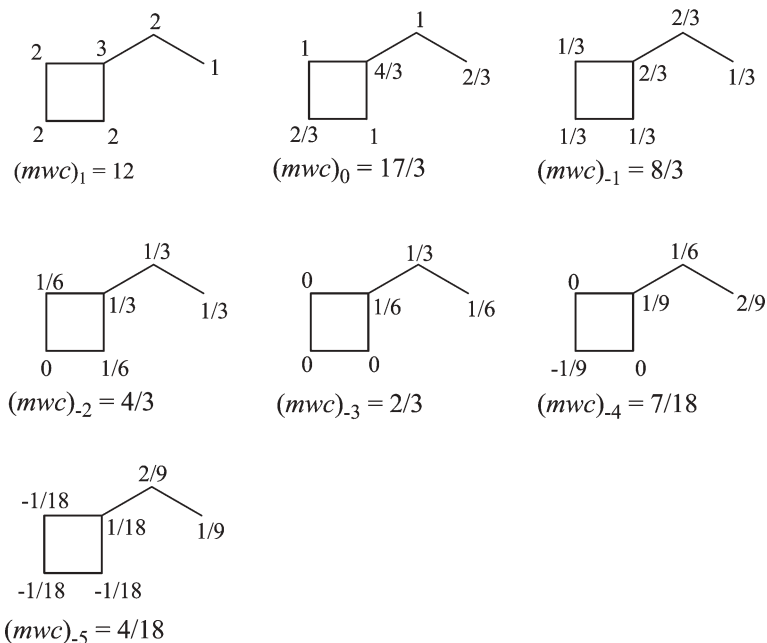


Figure 4 The values of $(awc)_\ell$ and $(mwc)_\ell$ for $\ell = 1, 0, -1, -2, -3, -4, -5$ for a graph representing the carbon skeleton of ethylcyclobutane

reviewed elsewhere^{182,215}) and for more general (chemical, biological, ecological) networks²¹⁶ – we consider here only those combinatorial quantities that have been used to characterize molecular complexity *via* the corresponding graphs. We already mentioned the role of walks in characterizing graph (molecular) complexity.¹⁸¹ Rucker and Rucker¹⁸¹ also introduced the term *labyrinthicity* if the complexity of a structure is measured by walk counts alone, because the walk count neglects symmetry. On the other hand, some other authors strongly incorporate symmetry characteristics of molecules. For example, Randić proposed a symmetry-dependent complexity measure based on the concept of augmented vertex-degree.^{217–219} The augmented degree of a vertex i in a (molecular) graph is the sum of its degree and degrees of all other vertices with weight $1/2^d$ depending on their distance d from the vertex i . The Randić complexity measure is then equal to the sum of the augmented degrees of vertices not equivalent by symmetry. It should be noted that the vertex-degrees have all kinds of nice combinatorial properties, *e.g.* refs. 220, 221, and they span the history of (chemical) graph theory from the beginnings (the Königsberg bridge problem, the handshaking lemma)²²² to our times, *e.g.* refs. 223–225. We should also remember that N -tuple code is based on vertex-degrees in a graph and its subgraphs.⁷ Some other authors have used in the past, for example, the degree distribution for the construction and enumeration of constitutional isomers in the alkane series.^{226,227}

Combinatorial properties that quite naturally correspond to a chemist's view of complexity expressed above are the number of kinds of connected subgraphs (substructures) N_s and the total number of connected subgraphs (substructures) N_t . A connected subgraph (substructure) is a subgraph (substructure) in which for every pair of vertices (atoms) there is a sequence of edges (bonds) connecting them. Connected subgraphs can be generated by removing one or more edges and/or vertices from a graph under the condition that what remains is in one piece. The graph itself is counted as a subgraph for formal reasons. In Figure 5 we give all subgraphs of a simple graph G representing the carbon skeleton of methylcyclobutane.

After the pioneering work of Gordon and Kennedy²²⁸ on the use of subgraphs in chemistry, Bertz and Herndon²²⁹ were the first to explore the idea of using subgraphs for measuring molecular complexity (and similarity). Bertz *et al.*^{230,231} were first to use N_s and N_t numbers to study molecular complexity and synthetic complexity. Bertz²³²⁻²³⁴ has also used these numbers to study the complexity of synthetic reactions and routes. Bonchev²³⁵ independently used the N_t number for assessing molecular complexity.

A problem with the subgraph-counting numbers N_s and N_t is that their growth is explosive. For example, these numbers for a relatively simple molecule such as cubane are¹⁸² 64 and 2441, respectively. Therefore, it is generally out of the question to construct all connected subgraphs of graph by hand. Naturally a computer program would be of use to enumerate all nonisomorphic connected subgraphs, and consequently to calculate N_s and N_t numbers, up to a certain graph (molecular) size. Such a program has been prepared by Rucker and Rucker²³⁶ and was named NIMSG (NonIsoMorphic SubGraphs). Their computer program first generates all connected subgraphs and then uses a combination of well-discriminating graph invariants to eliminate duplicates. Their program is applicable to simple graphs and to general graphs. Before this program, they prepared one that was applicable only to simple graphs (they called it SUBGRAPH).²³⁷ Since a connected subgraph can be described as a linear array of labels of adjacent edges, just as a path can be described as a linear array of labels of adjacent vertices, the connected subgraphs are generated by a path-finding algorithm working on the graph's edge-adjacency matrix. The *edge*-adjacency matrix differs from the *vertex*-adjacency matrix in considering adjacencies of edges instead of adjacencies of vertices.⁵ Rucker *et al.*²³⁸ studied the limits of the computer program NIMSG within the comprehensive graph samples which serve as supersets of the graphs corresponding to saturated hydrocarbons, both acyclic with up to 20 carbon atoms and (poly)cyclic with up to 10 carbon atoms. Since a fast computer method for reliably discriminating all nonisomorphic graphs is not available, Rucker *et al.*²³⁸ used in their work graph invariants of discriminating power as high as possible. They established that NIMSG, using as discriminatory graph invariants the combination of the Balaban index J^{239} and the eigenvalues of the distance matrix,²⁴⁰ is reliable to use within the domain of mono-, di-, tri-, and tetra-cyclic saturated hydrocarbon substructures with up to ten carbon atoms as well as of all alkane substructures with up to 19 carbon atoms.

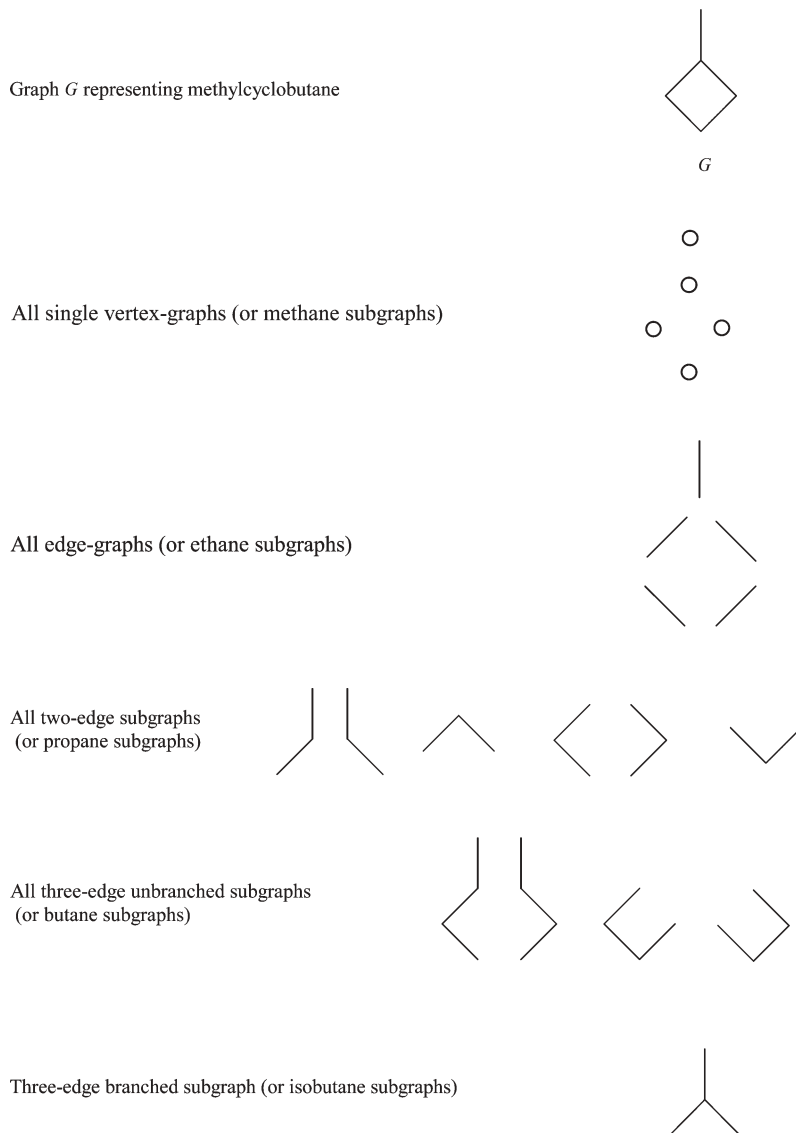


Figure 5 All connected subgraphs of a graph G representing the carbon skeleton methylcyclobutane

Therefore, the computer program NIMSG can be reasonably employed in chemistry whenever one is interested in computing substructures of a rather large number of saturated hydrocarbons which are themselves comprised of a modest number of vertices and edges.

When Rucker and Rucker¹⁹⁶ extended their program for the computation of N_s and N_t to unsaturated compounds represented by multigraphs, they had

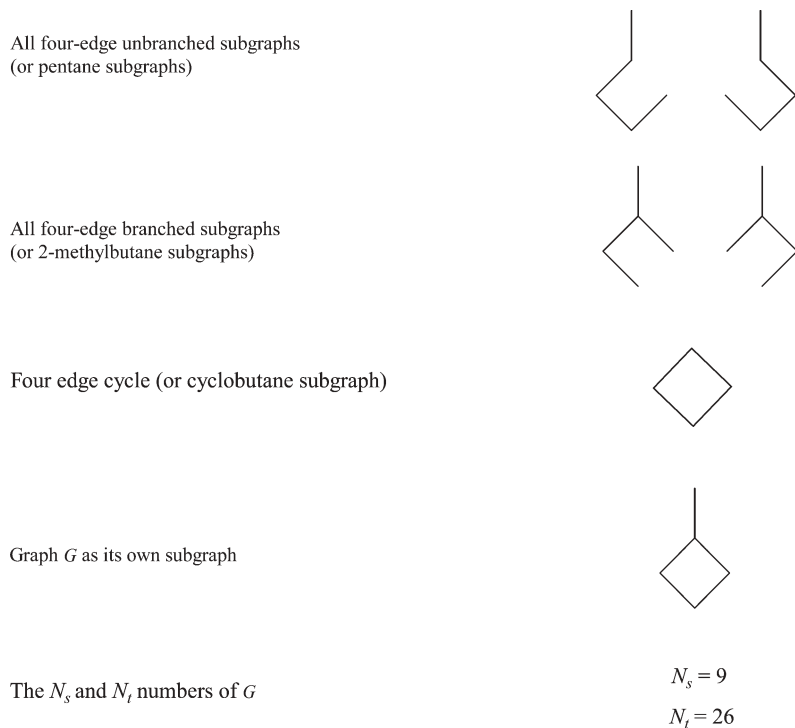


Figure 5 Continued

to distinguish between subgraph and substructure. The coincidence between subgraph and substructure, valid for simple graphs, does not hold for multigraphs. The difference between a substructure of an unsaturated molecule and a subgraph of the corresponding multigraph is that in the substructure all bonds still present have their multiplicities as they are in the parent molecule, whilst the subgraph may correspond to a less unsaturated analogue of the parent multigraph, that is, it may have a single or double edge where the multigraph has a double or triple edge. To illustrate this difference, we compare n -pentane and 1-pentene. In n -pentane (only the carbon skeleton is considered) one detects the following connected substructures: 5 methane substructures (single C atoms), 4 ethane substructures (C–C), 3 propane substructures (C–C–C), 2 butane substructures (C–C–C–C) and n -pentane (C–C–C–C–C) itself. Thus, the total number of connected substructures N_t is 15, whilst the number of distinct connected substructures N_s is 5. The total number of connected subgraphs N_t' and the number of distinct connected subgraphs N_s' are also 15 and 5, respectively, since n -pentane is a saturated molecule. In 1-pentene there are 5 methane substructures (single C atoms), 3 ethane substructures (C–C), 1 ethene substructure (C=C), 2 propane substructures (C–C–C), 1 propene substructure (C=C–C), 1 butane substructure (C–C–C–C), 1 butene (C=C–C–C) and 1-pentene itself. Hence, $N_t = 15$ and $N_s = 8$. The 5 methane

substructures give 5 single-vertex subgraphs; the 3 ethane and ethene substructures give 5 single-edge subgraphs; the ethene substructure gives additionally 1 double-edge 2-vertex subgraphs; the 2 propane and propene substructures give 4 three-vertex linear subgraphs; the propene substructure gives additionally 1 3-vertex multigraph; the butane and butene substructures give 3 four-vertex linear subgraphs; the butene substructure gives additionally 1 4-vertex multigraph and the 1-pentene graph gives 2 five-vertex linear subgraphs and its own subgraphs. Thus, $N_i' = 23$ and $N_s' = 9$. The N_i , N_s , N_i' and N_s' indicate that 1-pentene is more complex than n -pentane. Rucker and Rucker¹⁹⁶ designed a computer program to construct and count all connected subgraphs and substructures and distinct connected subgraphs and substructures of general graphs and the corresponding unsaturated heterocompounds. They named this program BERTZ, presumably in honour of Steven H. Bertz who in 1981 published a pioneering paper entitled *The First General Index of Molecular Complexity*²⁴¹ and has continued to contribute to the present day to our understanding of molecular complexity and complexity of chemical reactions.^{215,232-234} Program BERTZ is applicable to large sets of molecules, such as we find in combinatorial libraries, but in these cases because of the huge numbers of structures to process it is necessary to limit the maximal size of substructures (subgraphs) to a certain number of bonds (edges). In Figure 3 we also give N_i , N_s , N_i' and N_s' numbers for a set of molecules for which we listed their *twc* values – all the numbers are taken from Rucker and Rucker.¹⁹⁶

There have been some further miscellaneous uses of complexity measures. Arteca^{242,243} has discussed entanglement complexity for linear chain polymers. With the representation of a polymer conformation by the embedding of the linear-chain graph in 3-dimensions, a projection may be made into a 2-dimensional plane, whence the number of crossings there can be used as an entanglement complexity measure. Side-chain complexity enters into Davidson's lexicography³² for alkanes. And further in biology and in physics there has been work on various complexity measures.

The number of spanning trees is another classic²⁴⁴ combinatorial measure of graph complexity which has been used²⁴⁵ for (poly)cyclic molecules. A spanning tree of a (molecular) graph G is a connected, acyclic subgraph that includes all the vertices of G .^{4,5} Since acyclic graphs do not possess cycles, the spanning tree of an acyclic graph is identical to the graph itself. Therefore, the number of spanning trees can serve only as a measure of complexity in the case of (poly)cyclic graphs. Bonchev and coworkers²⁴⁶ were the first authors who used spanning trees (they called them maximal trees) explicitly as a measure of complexity for chemical structures, whilst Brooks *et al.*²⁴⁴ were first who defined the number of spanning trees of graphs as complexity for graphs. Independently, Gutman *et al.*²⁴⁵ used the number of spanning trees to study the complexity of cyclic graphs. There are several methods for obtaining the number of spanning trees.²⁴⁷ Here we will mention only the method based on the Laplacian matrix. The Laplacian matrix L of a graph G is defined as:²⁴⁸

$$L = D - A \quad (22)$$

The Laplacian matrix is sometimes also called the Kirchhoff matrix²⁴⁸ of a graph because of its role in the matrix-tree theorem,¹¹⁶ implicit²⁴⁹ in the work of Kirchhoff.²⁵⁰ And sometimes \mathbf{L} is called the *combinatorial* Laplacian, to distinguish it from the *normalized* Laplacian earlier noted in connection with random walks.

The Laplacian matrix is a real symmetric matrix, so that diagonalization of the Laplacian matrix of a graph (molecule) G with N vertices (atoms) gives N real eigenvalues λ_i , $i=1, \dots, N$. The smallest eigenvalue of the Laplacian spectrum $\lambda_i(\mathbf{L})$ is always 0, as a consequence of the special structure of the Laplacian matrix. The uses of the Laplacian matrix, its characteristic polynomial, its eigenspectrum, and related invariants have been explored in chemistry for at least the last decade.^{184,247,251–262}

The number of spanning trees, denoted by t , of a (molecular) graph is:^{248,263}

$$t = (1/N) \prod_{i=2}^N \lambda_i(\mathbf{L}) \quad (23)$$

In regular graphs (that is, graphs in which every vertex has the same degree r),⁴ the eigenvalues $\lambda_i(\mathbf{L})$ of the Laplacian matrix:

$$0 = \lambda_1(\mathbf{L}) < \lambda_2(\mathbf{L}) \leq \dots \leq \lambda_N(\mathbf{L}) \leq 2r \quad (24)$$

and of the vertex-adjacency matrix $\lambda_i(\mathbf{A})$:

$$r = \lambda_1(\mathbf{A}) < \lambda_2(\mathbf{A}) \leq \dots \leq \lambda_N(\mathbf{A}) \leq -r \quad (25)$$

are related by:¹¹⁶

$$\lambda_i(\mathbf{L}) = r - \lambda_i(\mathbf{A}) \quad (26)$$

In fullerenes whose graphs are regular graphs with $r=3$ (since all their vertices are of degree 3), eq. (23) transforms by means of eq. (26) into:²⁶⁴

$$t = (1/N) \prod_{i=2}^N [3 - \lambda_i(\mathbf{A})] \quad (27)$$

This formula was used by Fowler²⁶⁵ to study complexity of isomeric C_{40} , C_{60} and C_{70} fullerenes. He found that the most stable isomers of C_{40} , C_{60} and C_{70} fullerenes each have the greatest number of spanning trees and consequently the greatest complexity within their respective sets of 40, 1812 and 8149 candidates. Fowler²⁶⁵ obtained similar results when he applied the Randić complexity (RC) formula in the normalized form to the same three sets of isomeric fullerenes.^{217–219}

$$RC = (1/N) \sum_{i=1}^N \sum_{j=1}^N r_i / 2^{d_{ij}} \quad (28)$$

where r_i is the degree of the vertex i whilst d_{ij} is the distance in terms of the number of edges between vertices i and j . However, this formula produced less marked separation of isomers than eq. (27).

Mallion and Trinajstić²⁶⁶ used two combinatorial quantities *spanning-tree density* and *reciprocal spanning-tree density* in an attempt to quantify the structural intricateness (complicatedness) of (poly)cyclic molecular graphs. They referred to it by the term *intricacy*. The term *intricate* graphs has already been used by some other authors.²⁶⁷

The spanning-tree density of a simple (molecular) graph G , denoted by STD , is defined as:

$$STD = t / {}^E C_{N-1} \quad (29)$$

where ${}^E C_{N-1}$ is the number of ways of choosing any $N-1$ edges from the E available edges in G . STD represents the *probability* that if any set of $N-1$ edges in G is selected, and if the remnant set of $(E-N+1)$ edges in G is deleted, the resulting entity is a spanning tree.

In order to deal with numbers that are greater than 1, Mallion and Trinajstić²⁶⁶ introduced the reciprocal spanning-tree density, denoted by $RSTD$, as:

$$RSTD = {}^E C_{N-1} / t \quad (30)$$

The criterion of structural intricacy is – the bigger $RSTD$ is, the more intricate the structure is.

For some classes of graphs $RSTD$ can be given in closed form. For example, formulas for a complete graph K_N (that is, a regular graph G with N vertices and $r=N-1$) and a complete bipartite graph $K_{m,n}$ (that is, a complete graph which is made up of two sets of vertices – m and n – such that each vertex in the m set is joined to all and only to vertices in the n set; note that $N=m+n$) are given as:

$$RSTD(K_N) = (1/2)^{N(N-1)} C_{N-1} / N^{N-2} \quad (31)$$

$$RSTD(K_{m,n}) = {}^{mn} C_{m+n-1} / m^{n-1} \times n^{m-1} \quad (32)$$

Examples of the complete graph and the complete bipartite graphs are the *Kuratowski graph* K_5 and *Utilities graph* $K_{3,3}$, respectively. These graphs have been used to characterize planar graphs – a graph is planar if and only if it has *no* subgraph homeomorphic to K_5 or $K_{3,3}$.²⁶⁸ The Kuratowski graph was used, for example, in the study of the rearrangement of tetragonal-pyramidal complexes,²⁶⁹ whilst the Utility graph is grounded in a puzzle discussed by Dudeney in 1917 in his book *Amusements in Mathematics*²⁷⁰ in which he collected a

number of mathematically based puzzles. Dudeney presented a puzzle which can be visualized by the Utility graph under the title *Water, Gas and Electricity*, hence the name of the graph. The puzzle is to lay on water, gas and electricity pipes from three sources to each of three houses without any pipe crossing another. Also K_5 and $K_{3,3}$ were used in the discussion of the topological chirality of proteins.²⁷¹ A graph is topologically chiral if its embedding in Euclidean 3-space cannot be converted to its mirror image by a continuous deformation which avoids edge intersections. Nonplanarity is a necessary condition for topological chirality because a planar graph is achiral in 3-space.²⁷²

The intricacies of both K_5 ($RTSD=1.68$) and $K_{3,3}$ ($RSTD=1.56$) graphs are similar according to the $RSTD$ criterion. However, the corresponding numbers of spanning trees are rather different: $t(K_5)=125$ and $t(K_{3,3})=81$. The Kuratowski graph and the Utility graph are depicted in Figure 6.

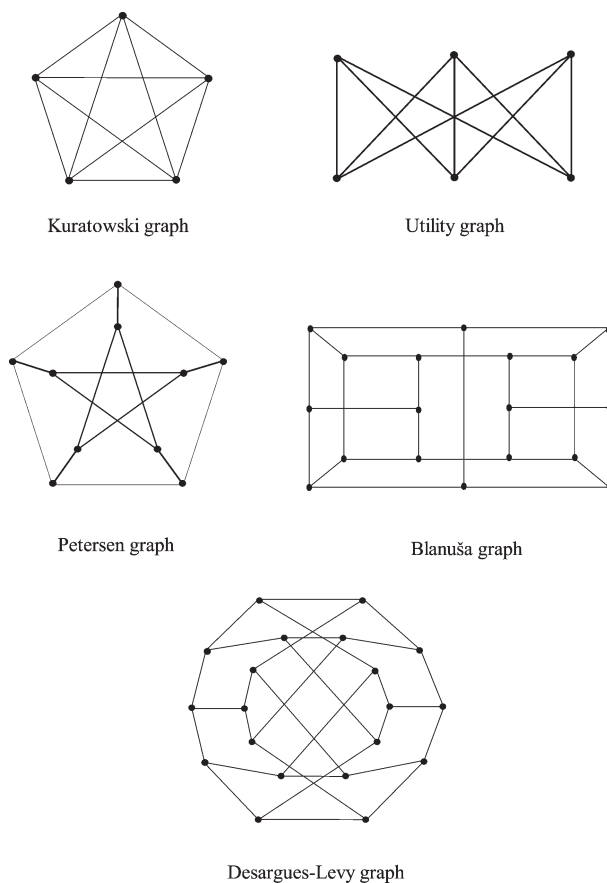


Figure 6 A collection of chemically-important graphs

Mallion and Trinajstić²⁶⁶ calculated the values of $RSTD$ for a number of interesting regular graphs of degree 3 which have found application in chemistry, such as: the Petersen graph²⁷³ ($RSTD=2.50$; $t=2000$) (which depicts possible routes for the isomerisation of trigonal-bipyramidal complexes²⁷⁴⁻²⁷⁶); the Blanuša graph²⁷⁷ ($RSTD=3.01$; $t=1037136$), which can be obtained by suitably combining two copies of the Petersen graph (the Blanuša graph has been introduced in the context of the four-color problem⁴); the Desargues-Levy graph²⁷⁸ ($RSTD=8.89$; $t=6144000$) (which has found application in describing isomerisations, rearrangements^{275,279-282} and nucleophilic displacement reactions²⁸³); and Schlegel graphs²⁸⁴ corresponding to C_{60} ($RSTD=3500$; $t=3.75 \times 1020$) and C_{70} ($RSTD=15000$; $t=1.14 \times 1024$) fullerenes (fullerenes probably being the most interesting class of chemical compounds which have emerged in the last two decades of the 20th century, when wide-spread research began on fullerene chemistry starting with the discovery of the first fullerene C_{60} , named buckminsterfullerene²⁸⁵). The Petersen graph, the Blanuša graph and the Desargues-Levy graph are also given in Figure 6, while C_{60} and C_{70} fullerenes and their Schlegel graphs are depicted in Figure 7.

Mallion and Trinajstić²⁶⁶ also considered Schlegel graphs of Platonic solids: the tetrahedron ($RSTD=1.25$; $t(16)$); the cube ($RSTD=2.06$; $t=384$); the octahedron ($RSTD=2.06$; $t=384$); the icosahedron ($RSTD=10.54$; $t=5184000$); and the dodecahedron ($RSTD=10.54$; $t=5184000$). A number of chemical structures possessing skeletons resembling Platonic solids are known, *e.g.* refs. 276, 286. The reason why the $RSTD$ - and t -values of the pair consisting of the

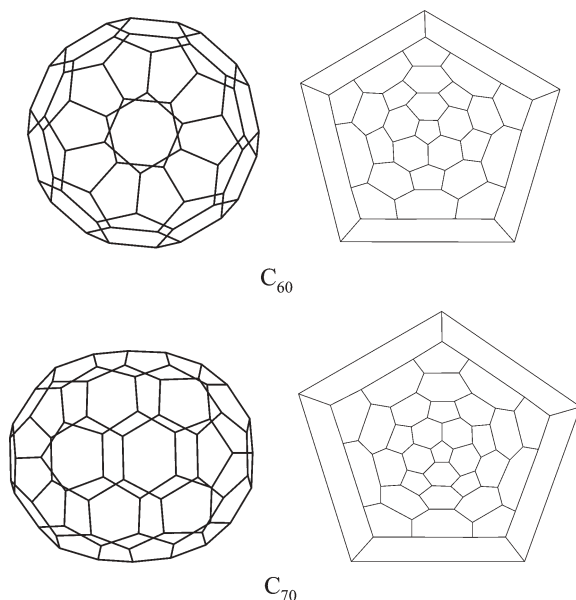


Figure 7 C_{60} and C_{70} fullerenes and their Schlegel graphs

cube and the octahedron as well as of the pair of the icosahedron and dodecahedron are the same is that each of these two pairs are duals⁴ to each other.²⁸⁷ The tetrahedron is its own dual.

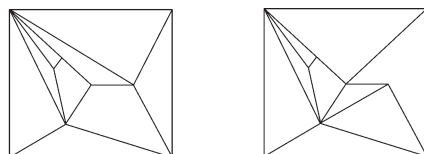
2.5 Other Enumerations. – Cash and Dias,²⁸⁸ in a continuation of Dias' work on developing methods for calculating eigenvectors, eigenvalues and resonance-structure counts for benzenoid radicals and diradicals, studied polyradicals. The resonance-structure counts of benzenoid polyradicals can be determined from the tail coefficients of their matching polynomials.²⁸⁹ Coefficients of the matching polynomial are made up from only acyclic contributions and this polynomial is a key concept in the topological resonance energy (TRE).^{290–292} The TRE theory is nowadays accepted by many as a reliable theory of aromaticity.²⁹³

Cash and Dias²⁸⁸ used *Mathematica*^{®294} to get the matching polynomials. In their calculation, they utilized a program by Salvador *et al.*²⁹⁵ and the fragmentation procedure in the case of large structures by Babić *et al.*²⁹⁶ The structure counts are large numbers for condensed benzenoid mono- and diradicals, but they become huge numbers in polyradicals. For example, the following are structure counts (given in brackets for only one diradical isomer of each formula) for the polycircumtriangulene diradical series: C₂₂H₁₂ (306), C₃₇H₁₈ (999), C₇₃H₂₁ (189792), C₁₂₁H₂₇ (153915460), whilst for the tetradical series: C₄₆H₁₈ (273956), C₈₈H₂₄ (604073400), C₁₄₂H₃₀ (3859415491248). Dias^{297,298} and Dias and Cash²⁹⁹ have made further resonance-structure counts for different classes of benzenoid radicals.

Hosoya *et al.*³⁰⁰ in continuation of their work on topological properties of polyhedral graphs, searched for twin graphs in polyhedral graphs with nine and ten vertices. Graphs corresponding to polyhedra are called polyhedral graphs.⁵ Thence polyhedral graphs are planar, and their planar representations are often termed Schlegel diagrams.²⁸⁴ Hosoya and co-workers³⁰¹ have previously defined topological twin graphs as pairs of highly similar nonisomorphic graphs that are isospectral with respect not only to the characteristic polynomial,^{5,52,116} but also to the distance polynomial,²⁴⁰ matching polynomial,²⁸⁹ the Z-counting polynomial³⁰² and have identical values of the Wiener index,³⁰³ Z-index³⁰² and many other topological indices.^{5,52,184,256}

The polynomial $P(x) = |\lambda I - A|$ of the vertex-adjacency matrix A of a graph G is the *characteristic* polynomial of G . The set of eigenvalues of A (*i.e.*, the zeros of $P(x)$) is called the *spectrum* of G . Isospectral graphs^{5,304–307} (sometimes also called cospectral graphs^{4,308}) are graphs with *identical* characteristic polynomials and consequently *identical* spectra. Hosoya and his co-workers^{300,301} prepared their own computer program to generate and enumerate polyhedral graphs. With this program they reproduced the known numbers of distinct polyhedral graphs with up to eleven vertices³⁰⁹ and obtained the number of polyhedral graphs with 12 vertices (6384636 graphs). In Hosoya *et al.*'s paper,³⁰⁰ isospectral pairing was checked for the polyhedral graphs with nine vertices (2606 graphs) and ten vertices (32300 graphs). Among 2606 polyhedral graphs with nine vertices they found 53 isospectral pairs, one isospectral triplet and one isospectral quartet. They are the smallest isospectral triplet and

quartet among the polyhedral graphs. Among 32300 polyhedral graphs with ten vertices they found many more isospectral pairs (622), nine isospectral triplets and two isospectral quartets. The isospectral pairs of polyhedral graphs with nine and ten vertices give rise to three and 15 pairs of topological twins. In Figure 8 we depict three pairs of the topological twin polyhedral graphs with nine vertices.

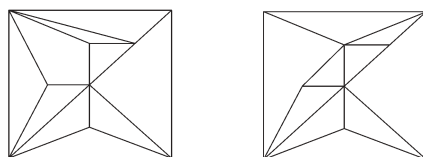


$$P(x) = x^9 - 17x^7 - 12x^6 + 58x^5 + 44x^4 - 56x^3 - 40x^2 + 12x + 8$$

$$S(x) = x^9 - 98x^7 - 616x^6 - 1502x^5 - 1422x^4 + 78x^3 + 834x^2 + 268x - 48$$

$$Q(x) = 1 + 17x + 84x^2 + 132x^3 + 46x^4$$

$$Z = 280 \quad W = 56$$

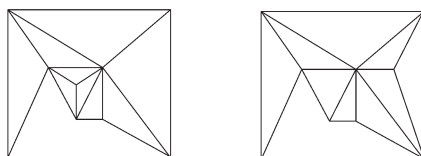


$$P(x) = x^9 - 18x^7 - 18x^6 + 66x^5 + 84x^4 - 55x^3 - 78x^2 + 12x + 18$$

$$S(x) = x^9 - 90x^7 - 554x^6 - 1378x^5 - 1544x^4 - 591x^3 + 146x^2 + 96x - 10$$

$$Q(x) = 1 + 18x + 94x^2 + 151x^3 + 52x^4$$

$$Z = 316 \quad W = 54$$



$$P(x) = x^9 - 18x^7 - 20x^6 + 62x^5 + 86x^4 - 58x^3 - 94x^2 + x + 12$$

$$S(x) = x^9 - 90x^7 - 560x^6 - 1426x^5 - 1694x^4 - 810x^3 + 10x^2 + 97x + 12$$

$$Q(x) = 1 + 18x + 92x^2 + 148x^3 + 53x^4$$

$$Z = 312 \quad W = 54$$

Figure 8 Schlegel diagrams of the three pairs of the topological twin polyhedral graphs with nine vertices taken from ref. 300. $P(x)$ denotes the characteristic polynomial, $S(x)$ the distance polynomial, $Q(x)$ the Z-counting polynomial, Z is the Hosoya Z-index and W the Wiener index

Rücker and Meringer³¹⁰ attempted to answer the following question: *How many organic compounds are graph-theoretically nonplanar?* The term *graph-theoretically nonplanar* means that the structural formula of such a compound cannot be drawn in a plane without at least two bonds crossing. It is interesting to note that most known compounds, even those that are geometrically quite clearly nonplanar such as helicenes, fullerenes and many other polycyclic compounds, correspond to planar graphs, *i.e.*, their structural formulas (considerably distorted, if necessary – see, for example, in Figure 7 Schlegel diagrams of C_{60} and C_{70} fullerenes and in Figure 8 Schlegel diagrams of six polyhedral graphs with nine vertices) can be drawn on a sheet of paper without two bonds crossing each other.

A justly famous graph-theoretical planarity criterion has been introduced and proved by Kuratowski.²⁶⁸ This states that a graph is non-planar iff it has no subgraph homeomorphic to either K_5 or $K_{3,3}$. However, graph-theoretical (*gt*) planarity tests available as algorithms or computer programs do not use the Kuratowski criterion – these tests are based on the trials to embed the structure in a plane without crossing of edges, a procedure more efficient than finding subgraphs of unknown structure.¹⁹⁵ (That is, there are many possible subgraphs which are homeomorphic to either K_5 or $K_{3,3}$.) Rücker and Meringer³¹⁰ used the Lempel-Even-Cederbaum planarity test³¹¹ written in C++ (a part of the Graph Template Library (GTL)³¹²). They first generated, using computer program MOLGEN 4.0,³¹³ all connected simple undirected graphs of nine vertices ($N=9$) and tested them for *gt*-(non)planarity in classes of constant edge-numbers E , that have also constant numbers of cycles $C (= E - N + 1)$. All graphs with $C \leq 3$ were found to be planar, all those of $C \geq 14$ were found to be nonplanar, and for $4 \leq C \leq 13$ the percentage of planar graphs among all such graphs monotonically decreases from 99.09 to 0.23%.

Next they performed a corresponding test on all simple 4-graphs (graphs with no vertex-degree higher than four – graphs that model carbon compounds) with $N=12$ vertices, in classes of constant E or C . And again, they found that all graphs with $C \leq 3$ were found to be planar, whilst for the classes with $4 \leq C \leq 13$ the percentage of planar graphs monotonically decreases from 98.80 to 0.84%.

Rücker and Meringer³¹⁰ performed two more tests – they considered all simple connected graphs with up to 12 vertices and all simple 4-graphs with up to 12 vertices. In the first case they found that beginning with $N=9$ the majority of simple connected graphs are nonplanar and in the second case that from $N=12$ the majority of simple 4-graphs are also nonplanar. Finally, Rücker and Meringer³¹⁰ considered all compounds in the MDL Beilstein file. Because of several technical, legal and financial obstacles, the *gt*-planarity test was carried out by D. Hounshell of MDL who ran the test on all Beilstein compounds in the MDL facilities (MDL, San Leandro, California). The *gt*-planarity test did not uncover compounds whose *gt*-nonplanarity was hitherto unknown. Rücker and Meringer³¹⁰ also tested peptides/proteins since it is known that *gt*-nonplanar peptides/proteins do exist.^{271,314–316} *gt*-Nonplanar peptides/proteins were retrieved from the CAS Registry file. They found that a very few proteins are *gt*-nonplanar.

Barone *et al.*³¹⁷ presented a program to generate a virtual library of structures. They named their program GASP (Génération Automatique de Structures Polycycliques). The authors claim that their program can generate all polycyclic structures with a given number of rings of given sizes. It works like the chemist who draws a ring, then adds another ring and so on. GASP starts with a ring, *i.e.*, its connection (connectivity) table,³¹⁸ then it adds the second ring in all possible positions, then, on each generated structure it adds the third ring and so on. The authors presented the results obtained for several structures (steroid, taxane and triquinane skeletons). For example, in the case of steroid-like skeleton (4 rings of size 6,6,6,5) GASP generates 988 polycyclic structures – among them there are 48 structures in which all 4 rings are fused. Comparison with the Beilstein database (<http://www.beilstein.com>) revealed that among these 48 structures there are 21 new structures. Barone *et al.*³¹⁷ were induced to publish GASP by a paper of Brinkmann *et al.*³¹⁹ in which these authors produced an algorithm for fast constructive enumeration of polycyclic chains with arbitrary ring sizes. The algorithm has been implemented as a computer program named *chains.c*. Computations have been carried out on Pentium II 350 Mhz. For example, there are 1681 polycyclic chains with 10 rings of size six (time needed to get this number was 0.02 seconds). If one has 10 rings such that five are of size 5 and five of size 6, then there are 391251 possible polycyclic chains (time needed for this number was 1.58 seconds).

3 Concluding Remarks

Combinatorial enumeration has played an important role in the history of chemistry.^{5,320–322} With the development and applications of combinatorial chemistry and (virtual and synthetic) combinatorial libraries catalysed by the development of powerful personal computers and portable software, combinatorial enumeration has become a chemist's, a biochemist's and a medicinal chemist's tool of everyday use.^{2,3,323} Therefore, it is beyond the point to discuss the usefulness of combinatorial enumeration in chemical, biochemical and medicinal research. Nevertheless, we wish to point to some recent achievements that were possible due to combinatorial enumeration.

Lahana and coworkers³²⁴ reported the rational design of immunosuppressive peptides without relying on information regarding their receptors or mechanisms of action. The design strategy was based on a virtual combinatorial library and the use of a variety of topological and shape descriptors with an analysis of molecular dynamics trajectories for the identification of potential drug candidates. Lahana *et al.*³²⁴ generated the virtual combinatorial library for the peptides of the general form RXXXRXXXXY. This sequence has seven positions to mutate in order to create the library. The use of 35 amino acids – 20 natural and 15 unnatural – leads to 35^7 combinations (64 billion compounds), well above the present computing power. Amino acids were characterized by their physical and chemical properties and also by their topological indices. The number of 35^7 compounds in the library was reduced

to 6^7 (279936) compounds by taking into account lipophilicity distribution, considered critical for the bioactivity studied. Screening the library of 279936 compounds, using two types of filters: static and dynamic, resulted in identification of 26 peptides satisfying all constraints. The bioactivity of these peptides was tested in a heterotopic mouse heart allograft model.³²⁵ The compound predicted to be the most potent displayed an immunosuppressive activity about 100 times higher than the lead compound. This kind of approach has become indispensable in drug design research, the first step being to build the combinatorial library.

Until about 1992 combinatorial libraries were almost entirely restricted to peptides and oligonucleotides,² but in the last decade they have been extended to many kinds of compounds.^{326,327} In this respect there is a very instructive recent article by Storm and Lünig³²⁸ on using virtual combinatorial libraries for the efficient preparation of macrocycles. However, combinatorial libraries are still predominantly used in drug research.^{323,329}

Important research related to combinatorial libraries concerns library screening.³³⁰ Virtual screening of a combinatorial library may accelerate drug lead discovery by selecting subsets of compounds according to their structural similarity or dissimilarity toward compound collections of specific bioactivity. One way to virtually screen the combinatorial library is by means of structural descriptors. The process of virtual screening of a combinatorial library starts from a wide selection of reactants that are used to generate *in silico* a huge number of compounds, according to multicomponent reaction.³³¹ In the next step, for each chemical compound a comprehensive set of structural descriptors is calculated, followed by a dimensionality reduction by selecting from the descriptors set a chemical space that is relevant for the investigated biological target. Finally, the compounds for synthesis and the high-throughput screening are selected with a statistical algorithm that implements a similarity, diversity or drug-like paradigm. There are many kinds of molecular descriptors in use:^{5,52,184,256,332–335} physicochemical or empirical (e.g., logarithm of octanol-water partition coefficient, $\log P$ ³³⁶), constitutional (e.g., number of hydrogen-bond donors, number of hydrogen-bond acceptors, number of rotatable bonds), graph invariants (e.g., cyclomatic number, walk counts), topological indices (e.g., Wiener index,³⁰³ Hosoya index,³⁰² Zagreb indices,^{337,338} Randić index,²⁰¹ Balaban index,²³⁹ Harary index^{339,340}), geometric (e.g., molecular volume, polar surface area), quantum (HOMO energy, atomic charges) and grid (various steric, electrostatic and lipophilic fields). Also there are rather general combinatorial questions concerning *data mining* and data manipulation in virtual combinatorial libraries. For a sampling of further recent references see.^{341–346}

Ivanciuc and Klein³⁴⁷ presented a simple and fast algorithm for building-block computation of a variety of Wiener-type indices for the virtual screening of very large combinatorial libraries. They short-cut the usual more detailed computational procedures for the indices for each individual compound. Also these two authors³⁴⁸ have reported the computation of Wiener-type indices for virtual combinatorial libraries generated from the heteroatom-containing

building-blocks. The approach is related to that used by Bytautas and Klein³⁴⁹ in computing average Wiener numbers for immensely large (*i.e.*, $>10^{24}$ structures). Ivanciuc³⁵⁰ reported an efficient algorithm for the building-block computation of the Ivanciuc-Balaban indices³⁵¹ for the virtual screening of any large combinatorial library.

We should end our report by pointing to the motivation as to why Vollhardt *et al.*³⁵² synthesized bent [4]phenylene using their own words: *This compound is of interest because it constitutes the fifth and last [4]phenylene isomer* (here they refer to the paper by Gutman *et al.*³⁵³ on the enumeration of the isomers of phenylenes), *and its preparation allows the completion of an experimental analysis of the comparative properties of this group of topomers, including data based on calculations.* [R]phenylenes (R =the number of benzene rigs) represent a subclass of [4N]annuleno-[4N + 2]annulenes in which (formally) benzene (aromatic) rings alternate with cyclobutadiene (anti-aromatic) rings.³⁵⁴ This kind of the structural arrangement of [R]phenylenes results in a number of interesting properties.³⁵⁵ Gutman *et al.*³⁵³ enumerated the [R]phenylene $C_{6R}H_{2R+4}$ isomers with up to $R=12$. There are five $C_{24}H_{12}$ isomers, 122 $C_{42}H_{18}$ isomers, and the number of phenylene isomers goes up quickly giving 101161 $C_{72}H_{28}$ isomers. The preparation of higher linear phenylenes would be of interest, among other things, to prove or disprove theoretical ideas³⁵⁶ about possibility that these compounds possess conducting properties.

Thence it seems that combinatorial enumeration in chemistry continues its historical tradition, with an increasing degree of activity, to an ever wider range of applications.

Acknowledgement

DB and NT acknowledge the support by the Ministry of Science and Technology of the Republic of Croatia (Grant Nos. 0098034 and 0098057) and DJK acknowledges the support of the Welch Foundation of Houston, Texas.

References

1. D.J. Klein, D. Babić and N. Trinajstić, in: *Chemical Modelling – Applications and Theory*, ed. A. Hinchliffe, Specialist Periodical Reports, The Royal Society of Chemistry, Cambridge, UK, 2002, Vol. 2, pp. 56–95.
2. W.A. Warr, in: *Encyclopedia of Computational Chemistry*, editor-in-chief P. von R. Schleyer, Wiley, Chichester, 1998, pp. 407–417.
3. *Handbook of Combinatorial Chemistry*, eds. K.C. Nicolaou, R. Hanks and W. Hartwig, Wiley-VCH, Weinheim, 2002.
4. F. Harary, *Graph Theory*, Addison-Wesley, Reading, MA, 2nd printing, 1971.
5. N. Trinajstić, *Chemical Graph Theory*, CRC Press, Boca Raton, FL, 2nd revised Edn., 1992.
6. J. von Knop, K. Szymanski, W.R. Müller, H.W. Kroto and N. Trinajstić, *J. Comput. Chem.*, 1987, **8**, 549–554.

7. J. von Knop, W.R. Müller, Ž. Jeričević and N. Trinajstić, *J. Chem. Inf. Comput. Sci.*, 1981, **21**, 91.
8. N. Trinajstić, S. Nikolić, J. von Knop, W.R. Müller, and K. Szymanski, *Computational Chemical Graph Theory: Characterization, Enumeration and Generation of Chemical Structures by Computer Methods*, Ellis Horwood Ltd., Simon & Schuster, Chichester, 1991.
9. I. Gutman, W. Linert, I. Lukovits and A.A. Dobrynin, *J. Chem. Inf. Comput. Chem.*, 1997, **37**, 349.
10. T. Došlić, *Croat. Chem. Acta*, 2002, **75**, 881.
11. A. Cayley, *Quart. J. Math.*, 1889, **23**, 376.
12. I. Lukovits and I. Gutman, *Croat. Chem. Acta*, 2002, **75**, 563.
13. I. Lukovits, *J. Chem. Inf. Comput. Sci.*, 1999, **39**, 563.
14. H.L. Morgan, *J. Chem. Doc.*, 1965, **5**, 107; A simplified interpretation of the Morgan indexing algorithm was devised in the following paper: V. Kvasnička and J. Pospichal, *J. Chem. Inf. Comput. Sci.*, 1990, **30**, 99.
15. I. Lukovits, *J. Chem. Inf. Comput. Sci.*, 2000, **40**, 361.
16. I. Lukovits, in: *Topology in Chemistry: Discrete Mathematics of Molecules*, eds. D.H. Rouvray and R.B. King, Horwood Publishing Ltd, Chichester, 2002, pp. 327–337.
17. D. Veljan, *Combinatorial and Discrete Mathematics*, Algorithm, Zagreb, 2001, p. 134.
18. D.I.A. Cohen, *Basic Techniques of Combinatorial Theory*, Wiley, New York, 1978, pp. 121–132.
19. R. Pauncz, *Alternant Molecular Orbital Method*, Saunders, Philadelphia, 1967.
20. T. Yamanouchi, *Proc. Physico-Math. Soc. Japan*, 1936, **18**, 436.
21. L. Pogliani, M. Randić and N. Trinajstić, *Int. J. Math. Educ. Sci. Technol.*, 2000, **31**, 811.
22. M. Randić, *J. Chem. Inf. Comput. Sci.*, 1986, **26**, 136.
23. M. Randić, *Croat. Chem. Acta*, 1986, **59**, 327.
24. M. Randić, S. Nikolić and N. Trinajstić, *J. Mol. Struct. (Theochem)*, 1988, **165**, 213.
25. S. Nikolić and N. Trinajstić, *Croat. Chem. Acta*, 1990, **63**, 155.
26. J. von Knop, W.R. Müller, K. Szymanski, S. Nikolić and N. Trinajstić, in: *Computational Chemical Graph Theory*, ed. D.H. Rouvray, Nova Sci. Publ., Commack, N.Y., 1990, pp. 9–32.
27. M. Randić, S. Nikolić and N. Trinajstić, *J. Chem. Inf. Comput. Sci.*, 1995, **35**, 357.
28. M.L. Contreras, J. Alvarez, M. Riveros, G. Arias and R. Rozas, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 964.
29. S. Davidson, *J. Chem. Inf. Comput. Sci.*, 1989, **29**, 151.
30. S. Davidson, *J. Chem. Inf. Comput. Sci.*, 1991, **31**, 417.
31. S. Davidson, *J. Chem. Inf. Comput. Sci.*, 1992, **32**, 215.
32. S. Davidson, *J. Chem. Inf. Comput. Sci.*, 2002, **42**, 147.
33. L. Bytautas and D.J. Klein, *J. Chem. Inf. Comput. Sci.*, 1998, **38**, 1063.
34. H.R. Henze and C.M. Blair, *J. Am. Chem. Soc.*, 1931, **53**, 3042.
35. H.R. Henze and C.M. Blair, *J. Am. Chem. Soc.*, 1931, **53**, 3077.
36. C.M. Blair and H.R. Henze, *J. Am. Chem. Soc.*, 1932, **54**, 1098.
37. C.M. Blair and H.R. Henze, *J. Am. Chem. Soc.*, 1932, **54**, 1538.
38. D.D. Coffman, C.M. Blair and H.R. Henze, *J. Am. Chem. Soc.*, 1933, **55**, 252.
39. H.R. Henze and C.M. Blair, *J. Am. Chem. Soc.*, 1933, **55**, 680.
40. H.R. Henze and C.M. Blair, *J. Am. Chem. Soc.*, 1934, **56**, 157.

41. C.-W. Lam, *J. Math. Chem.*, 1998, **23**, 421.
42. C.-W. Lam, *J. Math. Chem.*, 2000, **27**, 23.
43. C.-W. Lam, *J. Math. Chem.*, 2002, **31**, 333.
44. J. Wang, R. Li and S. Wang, *J. Math. Chem.*, 2003, **33**, 171.
45. J. von Knop, W.R. Müller, K. Szymanski and N. Trinajstić, *Computer Generation of Certain Classes of Molecules*, SKTH–Kemija u industriji, Zagreb, 1985, pp. 81–116.
46. I. Gutman and S.J. Cyvin, *Introduction to the Theory of Benzenoid Hydrocarbons*, Springer-Verlag, Berlin, 1989, pp. 33–49.
47. S.W. Golomb, *Polyominoes*, Scribner, New York, 1965.
48. W.F. Lunnon, in: *Graph Theory and Computing*, ed. R.C. Read, Academic Press, New York, 1972, pp. 87–100.
49. F. Harary, in: *Graph Theory and Theoretical Physics*, ed. F. Harary, Academic Press, London, 1967, pp. 1–41.
50. D.A. Klarner, *Can. J. Math.*, 1967, **19**, 851.
51. A.T. Balaban and F. Harary, *Tetrahedron*, 1968, **24**, 2505.
52. I. Gutman and O.E. Polansky, *Mathematical Concepts in Organic Chemistry*, Springer-Verlag, Berlin, 1986, pp. 59–62.
53. W.R. Müller, K. Szymanski, J. von Knop, S. Nikolić and N. Trinajstić, *J. Comput. Chem.*, 1990, **11**, 223.
54. N. Trinajstić, *J. Math. Chem.*, 1990, **5**, 171.
55. K. Balasubramanian, J.J. Kauffman, W.S. Koski and A.T. Balaban, *J. Comput. Chem.*, 1980, **1**, 149.
56. J. von Knop, K. Szymanski, Ž. Jeričević and N. Trinajstić, *J. Comput. Chem.*, 1983, **4**, 23.
57. J. von Knop, W.R. Müller, K. Szymanski, and N. Trinajstić, *J. Chem. Inf. Comput. Sci.*, 1990, **30**, 159.
58. D. Mašulović, R. Tošić, B.N. Cyvin and S.J. Cyvin, *MATCH Commun. Math. Comput. Chem.*, 1993, **29**, 165.
59. R. Tošić, D. Mašulović, I. Stojmenović, J. Brunvoll, B.N. Cyvin and S.J. Cyvin, *J. Chem. Inf. Comput. Sci.*, 1995, **35**, 181.
60. G. Caporossi and P. Hansen, *J. Chem. Inf. Comput. Sci.*, 1998, **38**, 610.
61. F. Chyzak, I. Gutman and P. Paule, *MATCH Commun. Math. Comput. Chem.*, 1999, **40**, 139.
62. G. Brinkmann, G. Caporossi and P. Hansen, *MATCH Commun. Math. Comput. Chem.*, 2001, **43**, 133.
63. G. Brinkmann, G. Caporossi and P. Hansen, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 842.
64. M. Vöge, A.J. Guttmann and I. Jensen, *J. Chem. Inf. Comput. Sci.*, 2002, **42**, 456.
65. I. Gutman, *Bull. Chem. Technol. Macedonia*, 2002, **21**, 53.
66. R.G. Harvey, *Polycyclic Aromatic Hydrocarbons*, Wiley-VCH, New York, 1997.
67. S. Hagen and H. Hopf, *Topics Curr. Chem.*, 1998, **196**, 47.
68. M.D. Watson, A. Fechtenkötter and K. Müllen, *Chem. Rev.*, 2001, **101**, 1267.
69. M. Randić, *Chem. Rev.*, 2003, **103**, 3449.
70. e.g., P. Phillips, *Nature*, 2000, **406**, 687.
71. Y. Sritana-Anant, T.J. Seiders and J.S. Siegel, *Topics Curr. Chem.*, 1998, **196**, 1.
72. D.J. Klein and A. Misra, *MATCH Commun. Math. Comput. Chem.*, 2002, **46**, 45.
73. F.H. Kohnke, A.M. Slavin, J.F. Stoddart and D.J. Williams, *Angew. Chem.*, 1987, **99**, 941.

74. K. Yamamoto, T. Harada, Y. Okamoto, H. Chikamatsu, M. Nakazaki, Y. Kai, T. Nakao, M. Tanaka, S. Harda and N. Kasai, *J. Am. Chem. Soc.*, 1988, **119**, 3578.
75. A. Borchardt, A. Fuchicello, K.V. Kilway, K.K. Baldrige and J.S. Siegel, *J. Am. Chem. Soc.*, 1992, **114**, 1921.
76. R.M. Cory and C.L. McPhail, *Tetrahedron Lett.*, 1996, **37**, 1987.
77. O. Kintzel, P. Luger, M. Weber and A.-D. Schluter, *Eur. J. Org. Chem.*, 1998, 99.
78. W.S. Burnside, *Theory of Groups of Finite Order*, University Press, Cambridge, 2nd Edn., 1911.
79. W. Hässelbarth, *Theoret. Chim. Acta*, 1984, **66**, 91.
80. C.A. Mead, *J. Am. Chem. Soc.*, 1987, **109**, 2130.
81. J. Brocas, M. Gielen and R. Willem, *The Permutational Approach to Dynamic Stereochemistry*, McGraw-Hill, New York, 1983.
82. S. Fujita, *Theoret. Chim. Acta*, 1989, **76**, 247.
83. S. Fujita, *Symmetry and Combinatorial Enumeration in Chemistry*, Springer-Verlag, Berlin, 1991.
84. S. El-Basil, *Combinatorial Organic Chemistry – An Educational Approach*, Nova Sci. Publ., Huntington, N.Y., 1999.
85. G. Pólya, *Acta Math.*, 1937, **68**, 145; translation of this classical paper of combinatorial analysis appeared in G. Pólya and R.C. Read, *Combinatorial Enumeration of Groups, Graphs and Chemical Compounds*, Springer-Verlag, New York, 1987.
86. S. El-Basil, *MATCH Commun. Math. Comput. Chem.*, 2002, **46**, 7.
87. H. Dolhaine and H. Hönic, *MATCH Commun. Math. Comput. Chem.*, 2002, **46**, 71.
88. H. Dolhaine and H. Hönic, *MATCH Commun. Math. Comput. Chem.*, 2002, **46**, 91.
89. M. van Almsick, H. Dolhaine and H. Hönic, *MATCH Commun. Math. Comput. Chem.*, 2001, **43**, 153.
90. T. Hudlicky, K.A. Abboud, D.A. Entwistle, R. Fan, R. Maurya, A.J. Thorpe, J. Balonick and B. Myers, *Synthesis*, 1996, 897.
91. T. Hudlicky, K.A. Abboud, J. Balonick, R. Maurya, M.L. Stanton and A.J. Thorpe, *Chem. Comm.*, 1996, 1717.
92. H. Dolhaine, H. Hönic and M. van Almsick, *MATCH Commun. Math. Comput. Chem.*, 1999, **39**, 21.
93. e.g., K. Drickamer and J. Carver, *Curr. Opin. Struct. Biol.*, 1992, **2**, 653.
94. A. Kerber, *MATCH Commun. Math. Comput. Chem.*, 2002, **46**, 151.
95. A. Kerber, *Applied Finite Group Action*, Springer-Verlag, Berlin, 1998.
96. W. Hässelbarth and A. Kerber, *MATCH Commun. Math. Comput. Chem.*, 2002, **46**, 199.
97. T. Carrell, E.A. Wintner, A. Bashir-Hashemi and J. Rebek, Jr., *Angew. Chem. Int. Ed. Engl.*, 1994, **33**, 2059.
98. T. Carrell, E.A. Wintner, A.J. Sutherland, J. Rebek, Jr., Y.M. Dunayevskiy and P. Vouros, *Chem. Biol.*, 1995, **2**, 171.
99. M. van Almsick, H. Dolhaine and H. Hönic, *J. Chem. Inf. Comput. Sci.*, 2000, **40**, 956.
100. W. Hässelbarth, *Theor. Chim. Acta*, 1985, **67**, 339.
101. R.M. Nemba and A.T. Balaban, *MATCH Commun. Math. Comput. Chem.*, 2002, **46**, 235.
102. R.W. Robinson, F. Harary and A.T. Balaban, *Tetrahedron*, 1976, **32**, 355.

103. S. Fujita and S. El-Basil, *MATCH Commun. Math. Comput. Chem.*, 2002, **46**, 121.
104. E.K. Lloyd, *MATCH Commun. Math. Comput. Chem.*, 2002, **46**, 215.
105. E.K. Lloyd and H. Dolhaine, *MATCH Commun. Math. Comput. Chem.*, 2002, **46**, 137.
106. S. Fujita, *Helv. Chim. Acta*, 2002, **85**, 2440.
107. R. Pauncz, *The Symmetric Group in Quantum Chemistry*, CRC Press, Boca Raton, FL, 1995.
108. S. Fujita, *J. Math. Chem.*, 2003, **33**, 113.
109. S. Fujita, *J. Am. Chem. Soc.*, 1990, **112**, 3390.
110. S. Cahn, C.K. Ingold and V. Prelog, *Angew. Chem. Int. Ed. Engl.*, 1966 **5**, 385.
111. S. Fujita, *Bull. Chem. Soc. Jpn.*, 2001, **74**, 1585.
112. S. Fujita and S. El-Basil, *J. Math. Chem.*, 2003, **33**, 255.
113. F. Zhang and W. Yan, *MATCH Commun. Math. Comput. Chem.*, 2002, **48**, 117.
114. W.C. Shiu, P.C.B. Lam, F. Zhang and H. Zhang, *J. Math. Chem.*, 2002, **31**, 405.
115. F. Torrens, *Int. J. Quantum Chem.*, 2002, **88**, 392.
116. D.M. Cvetković, M. Doob and H. Sachs, *Spectra of Graphs - Theory and Application*, 3rd revised & enlarged Edn., Barth-Verlag, Heidelberg, 1995.
117. L. Lovász and M.D. Plummer, *Matching Theory*, North-Holland, Amsterdam, 1986.
118. P.E. John, H. Sachs and M. Zheng, *J. Chem. Inf. Comput. Sci.*, 1995, **35**, 1019.
119. S.J. Cyvin and I. Gutman, *Kekulé Structures in Benzenoid Hydrocarbons*, Springer-Verlag, Berlin, 1988.
120. J.R. Dias, *Handbook of Polycyclic Hydrocarbons. Part A Benzenoid Hydrocarbons*, Elsevier, Amsterdam, 1987.
121. A. Misra and D.J. Klein, *J. Chem. Inf. Comput. Sci.*, 2002, **42**, 1171.
122. S.J. Cyvin, B.J. Cyvin, J. Brunvoll, H. Hosoya, F. Zhang, D.J. Klein, R. Chen and O.E. Polansky, *Monat. Chem.*, 1991, **122**, 435.
123. H.A. Staab and F. Diederich, *Chem. Ber.*, 1983, **116**, 3487.
124. D.J. Klein, T.G. Schmalz, G.E. Hite, A. Metropoulos and W.A. Seitz, *Chem. Phys. Lett.*, 1985, **120**, 367.
125. D.J. Klein, T.G. Schmalz and G.E. Hite, *J. Comput. Chem.*, 1986, **7**, 443.
126. D. Babić and A. Graovac, *Croat. Chem. Acta*, 1986, **59**, 731.
127. D.J. Klein, T.P. Živković and N. Trinajstić, *J. Math. Chem.*, 1987, **1**, 309.
128. A.N. Andriotis, M. Menon and D. Srivastava, *J. Chem. Phys.*, 2002, **117**, 2836.
129. G.M. Badger, *The Structure and Reactions of the Aromatic Compounds* University Press, Cambridge, 1954, pp. 89–90.
130. J.C. Dobrowolski, *J. Chem. Inf. Comput. Sci.*, 2002, **42**, 490.
131. J.K. Percus, *J. Math. Phys.*, 1969, **10**, 1881.
132. D. Cvetković, I. Gutman and N. Trinajstić, *Chem. Phys. Lett.*, 1972, **16**, 614.
133. D. Cvetković, I. Gutman and N. Trinajstić, *J. Chem. Phys.*, 1974, **61**, 2700.
134. N. Trinajstić, in: *Semiempirical Methods of Electronic Structure Calculation. Part A: Techniques*, ed. G.A. Segal, Plenum, New York, 1977, pp. 1–27.
135. H. Minc, *Permanents*, Addison-Wesley, Reading, MA, 1978.
136. G.G. Cash, *J. Math. Chem.*, 1995, **18**, 115.
137. I. Lukovits, A. Graovac, E. Kálmán, G. Kaptay, P. Nagy, S. Nikolić, J. Sytchev and N. Trinajstić, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 609.
138. S. Iijima, *Nature*, 1991, **354**, 5613.
139. R. Saito, G. Dresselhaus and M. Dresselhaus, *Physical Properties of Nanotubes*, Imperial College Press, London, 1998.

140. M. Terrones, W.-K. Hsu, H.W. Kroto and D.R.M. Walton, *Topics Curr. Chem.*, 1999, **199**, 189.
141. P. Avouris, *Acc. Chem. Res.*, 2002, **35**, 1026.
142. D.J. Klein, W.A. Seitz and T.G. Schmalz, *J. Phys. Chem.*, 1993, **97**, 1231.
143. G. Kaptay, I. Sytchev, J. Miklósi, P. Póczik, K. Papp, P. Nagy and E. Kálmán, in: *Advances in Molten Salts. From Structural Aspects to Waste Processing*, ed. M. Gaun-Escard, Begell, New York, 1999, pp. 257–262.
144. L. Gherghel, C. Kübel, G. Lieser, H.-J. Räder and K. Müllen, *J. Am. Chem. Soc.*, 2002, **124**, 13130.
145. C.N.R. Rao and A. Govindaraj, *Acc. Chem. Res.*, 2002, **35**, 998.
146. H. Dai, *Acc. Chem. Res.*, 2002, **35**, 1035.
147. L.P. Biro, Z.E. Horváth, L. Szalmas, K. Kertész, F. Wéber, G. Juhász, G. Radnóczy and J. Gyulai, *Chem. Phys. Lett.*, 2003, **372**, 399.
148. C. Velasco-Santos, A.L. Martínez-Hernández, A. Consultchi, R. Rodríguez and V.M. Castano, *Chem. Phys. Lett.*, 2003, **373**, 272.
149. A. Hirsch, *Angew. Chem. Int. Ed. Engl.*, 2002, **41**, 1853.
150. E.C. Kirby, *Croat. Chem. Acta*, 1993, **66**, 13.
151. H. Sachs, P. Hansen and M. Zheng, *MATCH Commun. Math. Comput. Chem.*, 1996, **33**, 169.
152. D.J. Klein and H. Zhu, *Discrete Appl. Math.*, 1996, **67**, 157.
153. P.E. John, *Croat. Chem. Acta*, 1998, **71**, 435.
154. D.J. Klein, T.P. Živković and R. Valenti, *Phys. Rev. B*, 1991, **43**, 723.
155. R. Swinborne-Sheldrake, W.C. Herndon and I. Gutman, *Tetrahedron Lett.*, 1975, 755.
156. W.A. Seitz, D.J. Klein, T.G. Schmalz and M.A. Garcia-Bach, *Chem. Phys. Lett.*, 1985, **115**, 139.
157. S.E. Stein and R.L. Brown, *Carbon*, 1985, **23**, 105.
158. S.E. Stein and R.L. Brown, *J. Am. Chem. Soc.*, 1987, **109**, 3721.
159. D.J. Klein, *Chem. Phys. Lett.*, 1994, **217**, 261.
160. D.J. Klein and L. Bytautas, *J. Phys. Chem.*, 1999, **93**, 5196.
161. O. Ivanciuc, L. Bytautas and D.J. Klein, *J. Chem. Phys.*, 2002, **116**, 4735.
162. O. Ivanciuc, D.J. Klein and L. Bytautas, *Carbon*, 2002, **40**, 2063.
163. M. Fujita, K. Wakabayashi, K. Nakada and K. Kusakabe, *J. Phys. Soc. Jpn*, 1996, **65**, 1920.
164. K. Wakabayashi, M. Fujita, K. Kusakabe and K. Nakada, *Czech. J. Phys.*, 1996, **46**, 1865.
165. K. Nakada, M. Fujita, B. Dresselhaus and M.S. Dresselhaus, *Phys. Rev. B*, 1996, **54**, 17954.
166. N. Tyutyulkov, G. Madjarova, F. Dietz and K. Müllen, *J. Phys. Chem. B*, 1998, **102**, 10183.
167. C.A. Coulson, M.A. Herraiz, M. Leal, E. Santos and S. Senet, *Proc. Roy. Soc. (London) A*, 1963, **274**, 461.
168. C.A. Coulson and M.D. Poole, *Carbon*, 1964, **3**, 275.
169. K.H. Lee, M. Causá and S.S. Park, *J. Phys. Chem. B*, 1998, **102**, 6020.
170. M. Hjort and S. Stafström, *Phys. Rev. B*, 2000, **61**, 14089.
171. P.L. Giunta and S.P. Kelty, *J. Chem. Phys.*, 2001, **114**, 1807.
172. J.R. Hahn, H. Kang, S. Song and I.C. Jeon, *Phys. Rev. B*, 1996, **53**, 1725.
173. F. Atamny, O. Spillecke and R. Schlögl, *Phys. Chem. – Chem. Phys.*, 1999, **1**, 4113.

174. S.M. Lee, Y.H. Lee, Y.G. Hwang, J.R. Hahn and H. Kang, *Phys. Rev. Lett.*, 1999, **82**, 217.
175. M. Randić, *J. Comput. Chem.*, 1980, **1**, 386.
176. M. Barysz and N. Trinajstić, *Int. J. Quantum Chem.: Quantum Chem. Symp.* 1984, **18**, 661.
177. M. Razinger, *Theoret. Chim. Acta*, 1986, **70**, 365.
178. M. Randić, M. Barysz, J. Nowakowski, S. Nikolić and N. Trinajstić, *J. Mol. Struct. (Theochem)*, 1989, **185**, 96.
179. A.S. Shalabi, *J. Chem. Inf. Comput. Sci.*, 1991, **31**, 483.
180. G. Rücker and C. Rücker, *J. Chem. Inf. Comput. Sci.*, 1993, **33**, 683.
181. G. Rücker and C. Rücker, *J. Chem. Inf. Comput. Sci.*, 2000, **40**, 99.
182. S. Nikolić, N. Trinajstić, I.M. Tolić, G. Rücker and C. Rücker, in: *Complexity in Chemistry – Introduction and Fundamentals*, eds. D. Bonchev and D.H. Rouvray, Taylor & Francis, London, 2003, pp. 29–89.
183. G. Rücker and C. Rücker, *J. Chem. Inf. Comput. Sci.*, 1999, **39**, 788.
184. R. Todeschini and V. Consonni, *Handbook of Molecular Descriptors*, Wiley-VCH, Weinheim, 2000, pp. 480–486.
185. R.A. Marcus, *J. Chem. Phys.*, 1965, **43**, 2643.
186. Ref. 17, p. 251.
187. I. Gutman and N. Trinajstić, *Chem. Phys. Lett.*, 1973, **20**, 257.
188. J. von Knop, W.R. Müller, K. Szymanski, M. Randić and N. Trinajstić, *Croat. Chem. Acta*, 1983, **56**, 405.
189. G. Rücker and C. Rücker, *J. Chem. Inf. Comput. Sci.*, 1991, **31**, 422.
190. C. Rücker and G. Rücker, *J. Math. Chem.*, 1992, **9**, 207.
191. O. Ivanciuc and A.T. Balaban, *Croat. Chem. Acta*, 1996, **69**, 63.
192. I. Gutman, C. Rücker and G. Rücker, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 739.
193. M. Razinger, *Theoret. Chim. Acta* 1982, **61**, 581.
194. F. Harary and A.J. Schwenk, *Pacific J. Math.*, 1979, **80**, 443.
195. C. Rücker and G. Rücker, *J. Chem. Inf. Comput. Sci.*, 1994, **34**, 534.
196. G. Rücker and C. Rücker, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 1457.
197. R.J. Wilson, *Introduction to Graph Theory*, Oliver & Boyd, Edinburgh, 1972, p. 10.
198. D.E. Johnson and J.R. Johnson, *Graph Theory With Engineering Applications*, Ronald Press, New York, 1972, p. 45.
199. I. Lukovits, A. Miličević, S. Nikolić and N. Trinajstić, *Internet Electronic J. Mol. Design*, 2002, **1**, 388; <http://www.biochempress.com>.
200. F.R.K. Chung, *Spectral Graph Theory*, American Mathematical Society, Providence, Rhode Island, 1997.
201. M. Randić, *J. Am. Chem. Soc.*, 1975, **97**, 6609.
202. D.J. Klein, *Croat. Chem. Acta*, 2002, **75**, 633.
203. M. Diudea, *J. Chem. Inf. Comput. Sci.*, 1996, **36**, 535.
204. M. Diudea, *J. Chem. Inf. Comput. Sci.*, 1996, **36**, 833.
205. A. Dress, S. Grünwald, I. Gutman, M. Lepović and D. Vidović, *MATCH Commun. Math. Comput. Chem.*, 2003, **48**, 63.
206. S. Grünwald, *Appl. Math. Lett.*, 2002, **15**, 1001.
207. A. Dress and I. Gutman, *Appl. Math. Lett.*, 2003, **16**, 389.
208. I. Lukovits and N. Trinajstić, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 1110.
209. J. Figueras, *J. Chem. Inf. Comput. Sci.*, 1993, **33**, 717.
210. G. Rücker and C. Rücker, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 1115
211. M. Mitchell Waldrop, *Complexity*, Touchstone/Simon & Schuster, New York, 1992.

212. D. Bonchev and W.A. Seitz, in: *Concepts in Chemistry: A Contemporary Challenge*, ed. D.H. Rouvray, Wiley, New York, 1997, pp. 353–381.
213. *Complexity in Chemistry – Introduction and Fundamentals*, eds. D. Bonchev and D.H. Rouvray, Taylor & Francis, London, 2003.
214. S. Nikolić, I. M. Tolić, N. Trinajstić and I. Baučić, *Croat. Chem. Acta*, 2000, **73**, 909.
215. S.H. Bertz, *Complexity in Chemistry: Introduction and Fundamentals*, eds D. Bonchev and D. Rouvray, Taylor & Francis, London, 2003, pp. 91–156.
216. D. Bonchev, *SAR QSAR Environ. Res.*, 2003, **14**, 199.
217. M. Randić, *Croat. Chem. Acta*, 2001, **74**, 683.
218. M. Randić and D. Plavšić, *Croat. Chem. Acta*, 2002, **75**, 107.
219. M. Randić and D. Plavšić, *Int. J. Quantum Chem.*, 2003, **91**, 20.
220. R.J. Cook, *J. Comb. Theory*, 1979, **B 26**, 337.
221. M. Trusczynski, *J. Graph Theory*, 1984, **8**, 171.
222. L. Euler, *Commentarii Academiae Scientiarum Imperialis Petropolitanae* 1736, **8**, 128; English translation appeared in *Sci. Am.*, 1953, **189**, 66.
223. M. Randić, *J. Chem. Inf. Comput. Sci.*, 2000, **40**, 627.
224. I. Lukovits, *MATCH Commun. Math. Comput. Chem.*, 2001, **44**, 279.
225. I. Lukovits, S. Nikolić and N. Trinajstić, *Chem. Phys. Lett.*, 2002, **354**, 417.
226. T.I. Bieber and M.D. Jackson, *J. Chem. Inf. Comput. Sci.*, 1993, **33**, 696.
227. M.D. Jackson and T.I. Bieber, *J. Chem. Inf. Comput. Sci.*, 1993, **33**, 701.
228. M. Gordon and J.W. Kennedy, *J. Chem. Soc. Faraday Trans. II*, 1973, **69**, 484.
229. S.H. Bertz and W.C. Herndon, in: *Artificial Intelligence Application in Chemistry*, eds. T.H. Pierce and B.A. Bohne, American Chemical Society, Washington, DC, 1986, pp. 169–175.
230. S.H. Bertz and W.F. Wright, *Graph Theory Notes of New York*, 1998, **35**, 32.
231. S.H. Bertz and T.J. Sommer, *Chem. Comm.*, 1997, 2409.
232. S.H. Bertz and C.M. Zamfirescu, *MATCH Commun. Math. Comput. Chem.*, 2000, **42**, 39.
233. S.H. Bertz, *New J. Chemistry*, 2003, **27**, 860.
234. S.H. Bertz, *New J. Chemistry*, 2003, **27**, 870.
235. D. Bonchev, *SAR QSAR Environ. Res.*, 1997, **7**, 23.
236. G. Rücker and C. Rücker, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 314; (Errata: *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 865).
237. G. Rücker and C. Rücker, *MATCH Commun. Math. Comput. Chem.*, 2000, **41**, 145.
238. C. Rücker, G. Rücker and M. Meringer, *J. Chem. Inf. Comput. Sci.*, 2002, **42**, 640.
239. A.T. Balaban, *Chem. Phys. Lett.*, 1982, **89**, 399.
240. Z. Mihalić, D. Veljan, D. Amić S. Nikolić, D. Plavšić and N. Trinajstić, *J. Math. Chem.*, 1992, **11**, 223.
241. S.H. Bertz, *J. Am. Chem. Soc.*, 1981, **103**, 3599.
242. G.A. Arteca, *J. Chem. Inf. Comput. Sci.*, 2002, **42**, 326.
243. G.A. Arteca, *J. Chem. Inf. Comput. Sci.*, 2002, **43**, 63.
244. R.L. Brooks, C.A.B. Smith, A.H. Stone and W.T. Tutte, *Duke Math. J.*, 1940, **7**, 312.
245. I. Gutman, R.B. Mallion and J.W. Essam, *Mol. Phys.*, 1983, **50**, 859.
246. D. Bonchev, O.N. Temkin and D. Kamenski, *React. Kinet. Catal. Lett.*, 1980, **15**, 119.
247. S. Nikolić, N. Trinajstić, A. Jurić, Z. Mihalić and G. Krilov, *Croat. Chem. Acta*, 1986, **69**, 883.

248. B. Mohar, in: *MATH/CHEM/COMP 1988*, ed. A. Graovac, Elsevier, Amsterdam, 1989, pp. 1–8.
249. J.W. Moon, *Counting Labelled Trees*, Canadian Mathematical Monographs, Ottawa, 1970, chapter 5.
250. G. Kirchhoff, *Ann. Phys. Chem.*, 1847, **72**, 497; English translation is given in J. B. O'Toole, *I.R.E. Trans. Circuit Theory*, 1958, **CT5**, 4.
251. O. Ivanciuc, *Rev. Roum. Chim.*, 1993, **38**, 1499.
252. B. Mohar, D. Babić and N. Trinajstić, *J. Chem. Inf. Comput. Sci.*, 1993, **33**, 153.
253. N. Trinajstić, D. Babić, S. Nikolić, D. Plavšić, D. Amić and Z. Mihalić, *J. Chem. Inf. Comput. Sci.*, 1994, **34**, 368.
254. D.J. Klein, *MATCH Commun. Math. Comput. Chem.*, 1997, **35**, 7.
255. D.J. Klein and H.-Y. Zhu, *J. Math. Chem.*, 1998, **23**, 179.
256. *Topological Indices and Related Descriptors in QSAR and QSPR*, eds. J. Devillers and A.T. Balaban, Gordon & Breach, Amsterdam, 1999.
257. I. Gutman and D. Vidović, *Indian J. Chem.*, 2002, **41A**, 893.
258. I. Gutman, *MATCH Commun. Math. Comput. Chem.*, 2003, **47**, 133.
259. I. Gutman, D. Vidović and B. Furtula, *Indian J. Chem.*, 2003, **42A**, 1272.
260. D.J. Klein and M. Randić, *J. Math. Chem.*, 1993, **12**, 81.
261. M. Kunz, *MATCH Commun. Math. Comput. Chem.*, 1995, **32**, 221.
262. B.E. Eichinger, *Macromol.*, 1980, **13**, 1.
263. B. Mohar, in: *Graph Theory, Combinatorics and Applications*, eds. Y. Alavi, G. Chartrand, O.R. Ollermann and A.J. Scwenk, Wiley, New York, pp. 871–898.
264. Z. Mihalić and N. Trinajstić, *Fullerene Sci. Technol.*, 1994, **2**, 89.
265. P.W. Fowler, *MATCH Commun. Math. Comput. Chem.*, 2003, **48**, 87.
266. R.B. Mallion and N. Trinajstić, *MATCH Commun. Math. Comput. Chem.*, 2003, **48**, 97.
267. e.g., G. Rücker and C. Rücker, *J. Chem. Inf. Comput. Sci.*, 1991, **31**, 123.
268. K. Kuratowski, *Fund. Math.*, 1930, **15**, 271.
269. M. Randić, V. Katović and N. Trinajstić, in: *Symmetries and Properties of Non-Rigid Molecules: A Comprehensive Survey*, eds. J. Maruani and J. Serre, Elsevier, Amsterdam, 1983, pp. 399–408.
270. H.E. Dudeney, *Amusements in Mathematics*, Nelson, London, 1917.
271. C. Liang and K. Mislow, *J. Am. Chem. Soc.*, 1994, **116**, 3588.
272. K. Mislow, *Croat. Chem. Acta*, 1996, **69**, 485.
273. J. Petersen, *L'Intermédiaire des Mathématiciens*, 1898, **5**, 225.
274. J.D. Dunitz and V. Prelog, *Angew. Chem.*, 1968, **80**, 700.
275. M. Randić, *Croat. Chem. Acta*, 1977, **49**, 643.
276. R.B. King, in: *From Chemical Topology to Three-Dimensional Geometry*, ed. A.T. Balaban, Plenum Press, New York, 1990, pp. 343–414.
277. D. Blanuša, *Glasnik mat. fiz. astr., ser 2*, 1946, **1**, 31.
278. H.S.M. Coxeter, *Bull. Am. Math. Soc.*, 1950, **56**, 413.
279. M. Gielen, in: *Chemical Applications of Graph Theory*, ed. A.T. Balaban, Academic Press, London, 1976, pp. 261–298.
280. M. Randić, D.J. Klein, V. Katović, D.O. Oakland, W.A. Seitz and A.T. Balaban, in: *Graph Theory and Topology in Chemistry*, eds. R.B. King and D.H. Rouvray, Elsevier, Amsterdam, 1987, pp. 266–284.
281. D.J. Klein, A. Graovac, Z. Mihalić and N. Trinajstić, *J. Mol. Struct. (Theochem)*, 1995, **341**, 157.
282. T. Živković, *Croat. Chem. Acta*, 1996, **69**, 215.
283. K. Mislow, *Acc. Chem. Res.*, 1970, **10**, 321.

284. H.S.M. Coxeter, *Regular Polytopes*, Dover, New York, 3rd Edn., 1973.
285. H.W. Kroto, J.R. Heath, S.C. O'Brien, R.F. Curl and R.E. Smalley, *Nature*, 1985, **318**, 162.
286. R.B. King, *Croat. Chem. Acta*, 2002, **75**, 447.
287. e.g., A. Rassat, *Angew. Chem.*, 2003, **115**, 632.
288. G.G. Cash and J.R. Dias, *J. Math. Chem.*, 2001, **30**, 429.
289. C.D. Godsil and I. Gutman, *J. Graph Theory*, 1981, **5**, 137.
290. I. Gutman, M. Milun and N. Trinajstić, *MATCH Commun. Math. Comput. Chem.*, 1975, **1**, 171.
291. J.-i. Aihara, *J. Am. Chem. Soc.*, 1976, **98**, 2750.
292. I. Gutman, M. Milun and N. Trinajstić, *J. Am. Chem. Soc.*, 1977, **99**, 1692.
293. e.g., V.I. Minkin, M.N. Glukhovtsev and B.Y. Simkin, *Aromaticity and Antiaromaticity*, Wiley-Interscience, New York, 1994, pp. 14–19.
294. *Mathematica*[®] is a registered trademark of Wolfram Research, Inc., <http://www.wolfram.com>
295. J.M. Salvador, A. Hernandez, A. Beltran, R. Duran and Mactuis, *J. Chem. Inf. Comput. Sci.*, 1998, **38**, 1105.
296. D. Babić, G. Brinkmann and A. Dress, *J. Chem. Inf. Comput. Sci.*, 1997, **37**, 920.
297. J.R. Dias, *J. Phys. Org. Chem.*, 2003, **15**, 94.
298. J.R. Dias, *J. Chem. Inf. Comput. Sci.*, 2001, **42**, 686.
299. J.R. Dias and G.G. Cash, *J. Chem. Inf. Comput. Sci.*, 2001, **41**, 129.
300. H. Hosoya, K. Ohta and M. Satomi, *MATCH Commun. Math. Comput. Chem.*, 2001, **44**, 183.
301. H. Hosoya, U. Nagashima and S. Hyugaji, *J. Chem. Inf. Comput. Sci.*, 1994, **34**, 428.
302. H. Hosoya, *Bull. Chem. Soc. Jpn*, 1971, **44**, 2332.
303. H. Wiener, *J. Am. Chem. Soc.*, 1947, **69**, 2636.
304. A.T. Balaban and F. Harary, *J. Chem. Doc.*, 1971, **11**, 258.
305. I. Gutman and N. Trinajstić, *Topics Curr. Chem.*, 1973, **42**, 49.
306. W.C. Herndon, *Tetrahedron Lett.*, 1974, 671.
307. T. Živković, N. Trinajstić and M. Randić, *Mol. Phys.*, 1975, **30**, 517.
308. V.L. Collatz and U. Sinogowitz, *Abh. Math. Sem. Univ. Hamburg*, 1957, **21**, 63.
309. R.C. Read and R.J. Wilson, *An Atlas of Graphs*, Oxford University Press, New York, 1998.
310. C. Rücker and M. Meringer, *MATCH Commun. Math. Comput. Chem.*, 2002, **45**, 153.
311. K.S. Booth and G.S. Lueker, *J. Comput. System Sci.*, 1976, **13**, 335.
312. M. Forstner, A. Pick and M. Raitner, *Graph Template Library GTL*, see <http://www.infosun.fmi.uni-passau.de/GTL/>.
313. A. Kerber, R. Laue, T. Grüner and M. Meringer, *MATCH Commun. Math. Comput. Chem.*, 1998, **37**, 205; see also R. Grund, A. Kerber and R. Laue, *MATCH Commun. Math. Comput. Chem.*, 1992, **27**, 87.
314. B. Mao, *J. Am. Chem. Soc.*, 1989, **111**, 6132.
315. C. Liang and K. Mislow, *J. Am. Chem. Soc.*, 1995, **117**, 4201.
316. K. Mislow and C. Liang, *Croat. Chem. Acta*, 1996, **69**, 1385.
317. René Barone, Remi Barone, M. Arbelot and M. Chanon, *Tetrahedron*, 2001, **57**, 6035.
318. M.F. Lynch, J.M. Harrison, W.G. Town and J.E. Ash, *Computer Handling Chemical Structure Information*, Macdonald, London, 1971.
319. G. Brinkmann, A.A. Dobrynin and A. Krause, *MATCH Commun. Math. Comput. Chem.*, 2000, **41**, 137.

320. D.H. Rouvray, *Chem. Soc. Rev.*, 1974, **3**, 355.
321. D.H. Rouvray, in: *Chemical Graph Theory – Introduction and Fundamentals*, eds. D. Bonchev and D.H. Rouvray, Gordon & Breach, New York, 1991, pp. 1–39.
322. A.T. Balaban, in: *Chemical Graph Theory – Introduction and Fundamentals*, eds. D. Bonchev and D.H. Rouvray, Gordon & Breach, New York, 1991, pp. 177–234.
323. A.E.P. Adang and P.H.H. Hermkens, *Curr. Med. Chem.*, 2001, **8**, 985.
324. G. Grassy, B. Calas, A. Yasri, R. Lahana, J. Woo, S. Iyear, M Kaczorek, R. Floc’h and R. Buelow, *Nature Biotechnol.*, 1998, **16**, 748.
325. R. Buelow, P. Veyron, C. Clayberger, P. Pouletty and J.L. Touraine, *Transplantation*, 1995, **59**, 455.
326. R.A. Lewis, S.D. Pickett and D.E. Clark, *Rev. Comput. Chem.*, 2000, **16**, 1.
327. A. Ivanisevic, K.V. McCumber and C.A. Mirkin, *J. Am. Chem. Soc.*, 2002, **124**, 11997.
328. O. Storm and U. Lüning, *Chem. Eur. J.*, 2002, **8**, 793.
329. T. Wright, V.J. Gillet, D.V.S. Green and S.D. Pickett, *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 381.
330. M.J. Blandamer, P.M. Cullis and P.T. Glesson, *Phys. Chem. Chem. Phys.*, 2002, **4**, 765.
331. R.W. Armstrong, A.P. Combs, P.A. Tempest, S.D. Brown and T.A. Keating, *Acc. Chem. Res.*, 1996, **29**, 123.
332. N. Trinajstić, M. Randić and D.J. Klein, *Acta Pharm.*, 1986, **36**, 267.
333. M. Randić, in: *Encyclopedia of Computational Chemistry*, Editor-in-chief P. von R. Schleyer, Wiley, Chichester, 1998, pp. 3018–3032.
334. M. Karelson, *Molecular Descriptors in QSAR/QSPR*, Wiley-Interscience, New York, 2000.
335. L. Poglianni, *Chem. Rev.*, 2000, **100**, 3827.
336. A. Leo, C. Hansch and D. Elkins, *Chem. Rev.*, 1971, **71**, 525.
337. I. Gutman and N. Trinajstić, *Chem. Phys. Lett.*, 1972, **17**, 535.
338. S. Nikolić, G. Kovačević, A. Miličević and N. Trinajstić, *Croat. Chem. Acta*, 2003, **76**, 113.
339. D. Plavšić, S. Nikolić, N. Trinajstić and Z. Mihalić, *J. Math. Chem.*, 1993, **12**, 235.
340. O. Ivanciuc, T.-S. Balaban and A.T. Balaban, *J. Math. Chem.*, 1993, **12**, 309.
341. V.J. Gillet, W. Khatib, P. Willett, P.J. Fleming and D.V.S. Green, *J. Chem. Inf. Comput. Sci.*, 2002, **42**, 375.
342. L. Xue and J. Bajorath, *J. Chem. Inf. Comput. Sci.*, 2002, **42**, 757.
343. A. Schuffenhauer, J. Zimmermann, R. Stoop, J.-J. van der Vyver, S. Lecchini and E. Jacoby, *J. Chem. Inf. Comput. Sci.*, 2002, **42**, 947.
344. S. Putta, C. Lemmen, P. Beroza and J. Greene, *J. Chem. Inf. Comput. Sci.*, 2002, **42**, 1230.
345. J.W. Raymond and P. Willet, *MATCH Commun. Math Comput. Chem.*, 2003, **48**, 197.
346. Y. Pan, N. Huang, S. Cho and A.D. MacKerell, Jr., *J. Chem. Inf. Comput. Sci.*, 2003, **43**, 267.
347. O. Ivanciuc and D.J. Klein, *Croat. Chem. Acta*, 2002, **75**, 577.
348. O. Ivanciuc and D.J. Klein, *J. Chem. Inf. Comput. Sci.*, 2002, **42**, 8.
349. L. Bytautas and D.J. Klein, *J. Chem. Inf. Comput. Sci.*, 2000, **40**, 471.
350. O. Ivanciuc, *Internet Electronic J. Mol. Design*, 2002, **1**, 1; <http://www.biochempress.com>.

351. O. Ivanciuc, T. Ivanciuc and A.T. Balaban, *J. Chem. Inf. Comput. Sci.*, 1998, **38**, 395.
352. D.T.-Y. Bong, L. Gentric, D. Holmes, A.J. Matzger, F. Scherbag and K.P.C. Vollhardt, *Chem. Comm.*, 2002, 278.
353. I. Gutman, S.J. Cyvin and J. Brunvoll, *Monat. Chem.*, 1994, **125**, 887.
354. B.C. Berris, G.H. Hovakeemian, Y.-H Lai, H. Mestdagh and K.P.C. Vollhardt, *J. Am. Chem. Soc.*, 1985, **107**, 5670.
355. C. Dosche, H.-G. Löhmansröben, A. Bieser, P.I. Dosa, S. Han, M. Iwamoto, A. Schleifenbaum, and K.P.C. Vollhardt, *Chem. Phys. Phys. Chem.*, 2002, **4**, 2156.
356. N. Trinajstić, T.G. Schmalz, T.P. Živković, S. Nikolić, G.E. Hite, D.J. Klein and W.A. Seitz, *New J. Chem.*, 1991, **15**, 27.