Internet Electronic Journal of
Molecular Design

July 2005, Volume 4, Number 7, Pages 434–490

Editor: Ovidiu Ivanciuc

Special issue dedicated to Professor Danail Bonchev on the occasion of the 65th birthday

My Life–Long Journey in Mathematical Chemistry

Danail Bonchev

Center for the Study of Biological Complexity, Virginia Commonwealth University, Richmond, Virginia 23284–2030

Received: February 24, 2005; Accepted: March 18, 2005; Published: July 31, 2005

Citation of the article:
D. Bonchev, My Life–Long Journey in Mathematical Chemistry, *Internet Electron. J. Mol. Des.* **2005**, *4*, 434–490, http://www.biochempress.com.

Inter*net* Electronic Journal of Molecular Design

BioChem Press
http://www.biochempress.com

My Life–Long Journey in Mathematical Chemistry

Danail Bonchev *

Center for the Study of Biological Complexity, Virginia Commonwealth University, Richmond, Virginia 23284–2030

Received: February 24, 2005; Accepted: March 18, 2005; Published: July 31, 2005

Internet Electron. J. Mol. Des. **2005**, *4* (7), 434–490

Abstract

The special issue of the *Internet Electron. J. Mol. Des.* represented a challenge to summarize the ideas and results, which I regard important in my 35 years research in the area of mathematical chemistry, and to look ahead to the paths mathematical chemistry might follow in the future. The importance of such an analysis stems from the generality of the methods used and their potential for a variety of applications. Topological patterns in molecular structures are identified and characterized quantitatively in four major classes: branching, cyclicity, centrality, and complexity. It is shown that it is through these patterns that topology controls the properties of molecules, polymers, and crystals. The direct relationships derived between the Wiener number, radius of gyration, and viscosity of polymers offer for the first time the chance to experimentally determine a topological index. Information theory was demonstrated as a method capable of capturing the essence of molecular and atomic structure, and to be a tool for investigating structure–property relationships. The predictions of the nuclear binding energies of nuclides of chemical elements 101–108, which we made 25 years ago, were recently confirmed with a relative standard deviation of only 0.1%. The overall connectivity indices and the information index for the vertex degree distribution were shown to be adequate measures of complexity of molecules and biological networks. Also reported are methods for classification, coding, enumeration and complexity assessment of chemical reaction networks. Mathematical chemistry is shown as a powerful tool for characterization of chemical structures with a variety of applications.

Keywords. Molecular topology; complexity; information theory; networks; kinetic graphs; atomic structure.

1 HOW IT ALL STARTED

Looking back through the crystal ball of time is nostalgic. You smile remembering your first steps in science, so clumsy and incommensurable to your great ambitions. You dig even further back and there they are – the roots of the real beginning, the two "chance" events that preprogrammed my future interests and development. I was only nine years old when a family relative, intrigued by my chess playing, presented me with a collection of chess magazines. This totally transformed my childhood and green years. Taking part in many competitions, I could play "blind" games, something that later turned into deriving equations in the same "blind" manner. Calculating chess moves made mathematics my favorite high school subject. Evaluating chess

---

* Correspondence author; phone: 804–827–7375; fax: 804–828–1961; E–mail: dgbonchev@vcu.edu.

positions taught me strategy and looking to a problem in its entirety, something that later proved to be an asset in planning and completing my research projects. A similar gift of Russian scientific–popular magazines, given to me by my uncle when I was 13, inspired a life–long interest toward science and science fiction. Growing up in a post–war communist country, in which there was no room for religion, it was science that became my religion. Science was the force that was moving mankind ahead, and where else could I be!

There was no room for a chance in selecting chemical engineering as my college major in the city of Sofia, the capital of Bulgaria. Several cousins of mine were already making university or industrial careers as chemists. Thus, the early combination of math and chemistry would inevitably one day bring me to mathematical chemistry, to which I devoted my life in science. However, that was still to come. After graduating from the technical university, I worked for three years as a process engineer in a big chemical plant. I was spending a good deal of my evening and night shifts with my favorite mathematical handbook. Another loved book was Brillouin's "Science and Information Theory" [1], which I managed to borrow for three months from the Library of Congress through our National Library. This fascinating book was my first contact with the "Big Science", a contact that turned Shannon's information theory [2] into a favorite research tool for the years to come.

Yet, it was still early for mathematical chemistry, and my D. Sc. degree in mathematical chemistry from the Moscow State University was light years away. The last two years in college awakened a deep interest toward the structure of atoms and molecules, and toward physics in general. It was then quite natural to accept the next generous gift of fate – the offer to found the department of physical chemistry at the newly opened technical university in my native town of Bourgas, on the Black Sea coast. Organizing laboratories and classes was combined with getting (externally) a Ph. D. degree in quantum chemistry under the guidance of Prof. Tyutyulkov from the Bulgarian Academy of Sciences. While specializing for six months in the Department of Quantum Mechanics at the University of St. Petersburg, Russia, I had the rare chance to extend my knowledge of physics, spending nights with Einstein's relativity theory and the physics of discrete space–time. Back to Bulgaria, the initial enthusiasm about quantum chemistry soon had to face the hard reality of making science in a small East–block country lacking the big computers of the leading Western countries. The Copenhagen interpretation of quantum mechanics, which I found unsatisfying, added to the accumulating disappointment. Lacking a college degree in physics, I had to abandon my dreams of a new physical theory, dreams that at the age of 27 sent me to the National Conference in Physics to report my information theory–based hypothesis for the dependence of the fine structure constant, mass and charge of electron on the age of the Universe. Thus, in the beginning of the 1970s I was ready for a change. I was ready for mathematical chemistry!

In fact, without knowing it, I had already been a mathematical chemist, applying information theory for a number of years to the characterization of atoms and molecules. Lacking self–confidence, I was keeping the unpublished results on my desk. Until one day Nenad Trinajstić, who was visiting me, looked at them and asked: "Man, why don't you publish all of this?" That was the encouragement, so much needed in the beginning, the turning point after which the earth started revolving much faster. In the next year, in January 1976, I returned his visit and spent three months in the beautiful city of Zagreb. I was not totally absorbed by the initiated work on applying my information theoretic indices to QSAR of a series of molecules, so I started playing with the chemical structures. I was assigning numbers to atoms and counting how many bonds separate them. One day, Ivan Gutman, who worked in the same room, looked over my shoulder and said: "Congratulations, Sir! You have reinvented the Wiener number." Then, he showed me the Wiener papers [3] and that of Hosoya [4] of 1971, and that was how I first met with graph theory, my great love. At that time, the Zagreb group of theoretical chemists (Trinajstić, Gutman, Graovac, Živković) was already very active and gained a lot of attention with the their work on the interplay between molecular orbital theory and graph theory [5]. My three months visit gave birth to our JCP paper with Nenad on molecular branching [6], in which the marriage of information theory and graph theory produced for the first time a detailed characteristic of this general topological property of molecular structures. Three years later, during the 1979 Bremen symposium in Germany, I was already proposing the creation of an International Society for Mathematical Chemistry.

This is how it all started. I was blessed to join the newly born area of mathematical chemistry during the time it was taking shape and was struggling for its place as a branch of theoretical chemistry. In the years to follow I was privileged to have contacts, friendships, and/or collaborations with most of the scientists who shared the same passion toward mathematical chemistry. All of the work performed until 1992, when I moved to the United States, would be impossible without the invaluable help of my former students/coworkers from the Bourgas Technical University "Prof. Assen Zlatarov": Verginia and Dimitar Kamenski, Bresitsa Rousseva, Todor Peev, Dimcho Dimov, Kristiana Tashkova, and, first of all, my former student, collaborator and close friend Professor Ovanes Mekenyan, with whom I founded in 1984 the Bourgas Laboratory of Mathematical Chemistry and Chemical Informatics, which he brilliantly developed as a world–class laboratory for QSAR software in environmental research and drug design. This article, which attempts to summarize some of my contributions to mathematical chemistry, is dedicated to the wonderful community of mathematical chemists, which I have had the privilege to know and collaborate with. Due to the scope and volume of the article, the references given are anything but complete, and the readers are addressed for more references to the original papers cited.

2 INFORMATION THEORY IN DESCRIBING ATOMS AND ATOMIC NUCLEI

2.1 My First Information Theory Study

One of the simplest formulas in Shannon's Information Theory [1,2] is the one describing the amount of information obtained in an "experiment" reducing the uncertainty of the possible *equiprobable* outcomes from P_0 to P_1:

$$I = \log_2 \frac{P_0}{P_1},\ bits \tag{2.1}$$

I made use of this formula [7] in 1960s in searching the answer of an important question: "What is the maximum amount of information that can be obtained in measuring distance and time?" Such a question has an answer only if space and time are discrete and finite. Assuming the radius R_o and age T_o of the Universe as an upper limit of the distance and time scale, while taking the classical radius of electron r_o and the time t_o for which light passes through electron as lower limits of the two scales, respectively, one obtains

$$I = \log_2 \frac{R_0}{r_0} = \log_2 \frac{T_0}{t_0} = 136.9 \quad bits \tag{2.2}$$

(Here, the logarithm at a base 2 is used to obtain information in binary digits, bits). The numerical result shown in Eq. (2.2) was obtained assuming the age of the Universe to be 5×10^{10} years, which was taken as an average of the 1960s estimates 1.3×10^{10} to 10^{11} years. This result coincided with an error of less than 0.1% with the reciprocal value of the important dimensionless fine structure constant $\alpha = 2\pi e^2/hc$, which incorporates three fundamental constants – the Planck constant h, velocity of light c, and electron charge e. Proceeding from the hypothesis that the numerical coincidence is in fact a real relationship, one concludes that the fine structure constant is in fact a variable quantity that depends on the Universe age:

$$\frac{1}{\alpha} = \frac{hc}{2\pi e^2} \approx \log_2 \frac{T}{t_0} \approx \log_2 \frac{R}{r_o} = 137.0\ bits \tag{2.3}$$

Thus, the dimensionless constant α was hypothesized to have a dimension of information or even to represent a conversion factor between a bit of information and a quantum of distance or time. It was also assessed from the equation that the annual variation of the fine structure constant would be about 1.9×10^{-11} %, and that the largest part of the Universe evolution would occur in a very small fraction of a second, in agreement with the Big Bang theory. Eq. (2.3) also indicated that at least one of the three fundamental constant included would also "age" with the Universe. Assuming this to be the electron charge, one easily arrives to conclusions and quantitative estimates for the Universe aging being associated with a decrease in electron charge, mass, and velocity, an increase in the atomic radius and that of the earth, as well as with atomic spectra red shift.

Papers relating the cosmic “big numbers” to the parameters of the microworld have been traditionally published in *Nature* [8,9]. Indeed, it was naïve to think that a manuscript coming from beyond the iron curtain, written in bad English by an unknown author, might also be published in this most respectful journal. Yet, it hurts when your paper is rejected on the ridiculous ground that your results have been previously published, and the reference given is totally irrelevant. Thus, this interesting hypothesis had to be buried in the Yearbook of my university [7]. In the years that followed, my “youth dreams” were popping–up spontaneously from time to time. They have left their traces in the wild idea to connect the upper end of the Periodic Table of chemical elements to the Eddington Number (the total number of particles in the universe) [10], as well as in my “spiraling electron” theory (a manuscript [11] summarized mainly during my ten days stay in the Hidden Valley Yoga Retreat in Southern California in the year 1997, and never submitted for publication). However, the rejection of my paper by *Nature* in the late sixties was the turning point after which, for good or for bad, I turned from physics to chemistry, and to the chemical applications of information theory and graph theory.

2.2 Information Theory As Applied To Discrete Finite Structures

In characterizing atoms and molecules one cannot proceed from Eq. (2.1), because the probabilities of different “events” are in general not equal and one deals with *probability distributions*. A convenient basis in these cases is the finite probability scheme proposed by Mowshowitz [12]. Let a system be composed of N elements, distributed according to certain equivalence criterion α into k groups, having N_1, N_2,…, N_k elements, respectively. A probability $p_i = N_i / N$ can be assigned to a randomly chosen element i to belong to the i^{th} group. A probability distribution $P(\alpha) = \{p_1, p_2, \ldots, p_k\}$ is thus constructed. The mean and total *entropy* of this distribution, ($\overline{H}(\alpha)$ *and* $H(\alpha)$, respectively) can be calculated by the Shannon formulas:

$$\overline{H}(\alpha) = -\sum_{i=1}^{k} p_i \log_2 p_i, \quad \textit{bits per element} \tag{2.4}$$

$$H(\alpha) = N \log_2 N - \sum_{i=1}^{k} N_i \log_2 N_i, \quad \textit{bits} \tag{2.5}$$

When all probabilities are equal, the entropy of the system is maximal:

$$\overline{H}_{\max}(\alpha) = \log_2 N; \quad H_{\max}(\alpha) = N \log_2 N \tag{2.6 a,b}$$

Any system that deviates from the state of maximum entropy is regarded to contain information or to have *information content*. The information content $I(\alpha)$ of a system is thus defined by the difference between the maximum possible entropy of a system with the same number of elements and the actual entropy of the system:

$$I(\alpha) = H_{\max}(\alpha) - H(\alpha) \tag{2.7}$$

Depending on the selected equivalence criterion α, a system can be characterized by different types of entropy or information content.

2.3 Information Content of a Chemical Element and a Nuclide [13]

In defining the information content of atoms we proceeded first by the distribution of the protons, neutrons and electrons into two substructures: atomic nucleus and electron shell. The *information content of a nuclide* having z protons, z electrons and n neutrons (*i.e.*, having atomic mass $A = z + n$) was defined as:

$$I_{nuclide} = (A+z)\log_2(A+z) - A\log_2 A - z\log_2 z \tag{2.8}$$

We then defined the information content of a chemical element as the average information content of its naturally occurring nuclides i, accounting for their abundance c_i:

$$I_{chem.element} = \sum\nolimits_i c_i I_{nuclide,i} \tag{2.9}$$

The properties of this information index remained largely unexplored except the few correlations with atomic properties reported [13,14]. Instead, the focus of our studies was shifted to the distributions of electrons in electronic shells and to the protons and neutrons distribution in the atomic nuclei. Proceeding from the quantum mechanical description of atom, the following information indices were defined [15–18]: information on the distribution of electrons over atomic shells, I_n, subshells, I_{nl}, atomic orbitals, I_{nlm}, and spin–orbitals, I_{nlmms}, as well as information on the electron distribution over the values of the orbital, magnetic, and spin quantum numbers, I_l, I_m, and I_{ms}. In all these cases, one deals with the distribution of z electrons into k groups, having z_1, z_2, …, z_k electrons, respectively. Denoting by x the combination of quantum numbers (the Shannon equivalence criterion α), which determine the type of distribution, we defined the respective total information content, $I(x)$, and the mean information content, $\bar{I}(x)$, as

$$I(x) = z\log_2 z - \sum\nolimits_i z_i \log_2 z_i \tag{2.10}$$

and

$$\bar{I}(x) = -\sum\nolimits_i \frac{z_i}{z}\log_2\frac{z_i}{z} \tag{2.11}$$

respectively.

A third information function termed *differential information*, $\Delta I_x(z)$, was also defined as the increase in the atomic information content $I_x(z)$ in a chemical element with atomic number z and combination of atomic quantum numbers x, as compared to that of the element having atomic number z–1:

$$\Delta I_x = I_x(z) - I_x(z-1) \tag{2.12}$$

The filling of a given electron subset (shell, subshell, etc.) *k*+1 begins and ends most frequently at a constant population of the preceding *k* subsets. Then, eq (2.12) may be transformed into

$$\Delta I_x = z\log_2 z - (z-1)\log_2(z-1) - (z_{k+1}\log_2 z_{k+1} - (z_{k+1}-1)\log_2(z_{k+1}-1)) \tag{2.13}$$

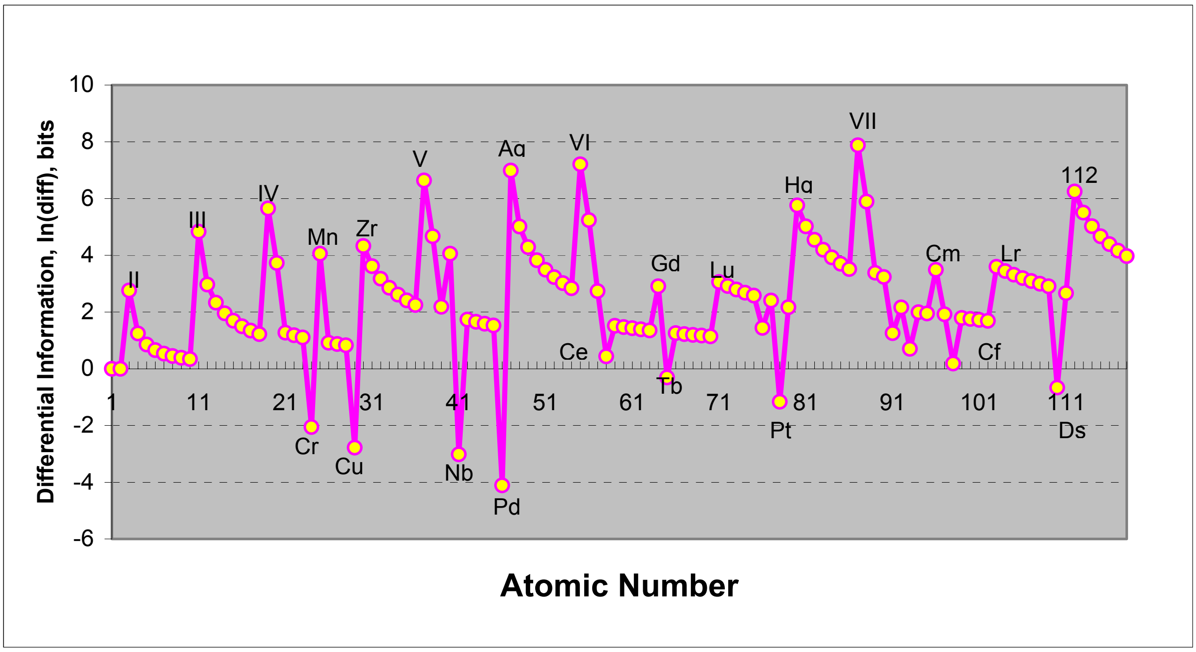


Figure 1. Differential information on the distribution of atomic electrons over electron shells.

The properties of the differential information function (2.13) enable the characterization of periodicity of chemical elements by a sharp maximum in the first element of each period or s–, p–, d–, and f– subperiod, followed by a gradual decrease to a minimum in the corresponding last element, as shown in Figure 1. This regular trend is best demonstrated by periods II and III. In periods IV to VII, the filling of the *d*– and *f*–subshells with delay makes the increase in the I_n index upon adding *d*–electrons lesser than that upon adding *p*–electrons, and the corresponding increase for the *f*–electrons lesser than that for the *d*–electrons:

$$\Delta I_n\,((\text{n}-2)\text{f}) < \Delta I_n\,((\text{n}-1)\text{d}) < \Delta I_n\,(\text{np}) < \Delta I_n\,(\text{ns}) \tag{2.14}$$

The violations in the "ideal order" of filling electron *f*– and *d*–subshells, caused by the accelerated adding of (*n*–1)*d*– or (*n*–2)*f*–electrons at the cost of the decreased population of the *ns*–subshell is also mirrored by the information function by a sharp minimum for the extra *d*–electron in Cr, Cu, Nb, Pd, etc., and by a maximum for restoring the "normal" trend in Mn, Zn, Ag, etc. The minimum in the *f*–elements, appears in the element returning the extra *d*–electron from the preceding element(s) back to the *s*–subshell: Tb after Gd, Np after Pa and U, Cf after Cm and Bk. The minima and maxima for the "anomalous" *f*–elements are lower than those for the corresponding *d*–elements, following the ordering, given above in inequalities (2.14).

2.4 Derived Information Equations For the Periods and Groups In the Periodic Table [19,20]

The first equation of this type, Eq. (2.15), contains the period constant A in which is equal for periods I through VII to 0, 2, 19.51, 37.02, 87.75, 138.48, and 242.51 bits, respectively. The group constants k_{nl} = 1 or 2 for groups 1 and 2, k_{nl} = 1 to 10 for groups 3 to 12, and k_{nl} = 1 to 6 for groups 13 to 18, respectively:

$$I_{nl} = z\log_2 z - A - \sum_l k_{nl}\log_2 k_{nl} \tag{2.15}$$

The total information on electron distribution over atomic orbitals, expressed in bits per atom, is another equation for the periods and groups in the Periodic Table:

$$I_{AO} = (z_0 + a + b)\log_2(z_0 + a + b) - z_0 - b \tag{2.15}$$

Here, b is the number of paired electrons in the atomic orbitals of the element. In the ground state of the atoms of elements of groups 1, 2, and 13 to 18, the group constant a is equal to the lowest valence of the element (1, 0, 1, 2, 3, 2, 1, 0), whereas in the highest valence state it equals the old numbering of the eight main groups (1 to 8). The period constant z_0 is the atomic number of the noble gas that ends the preceding period (z_0 = 2, 10, 18, 36, 54, 86).

The information on electron distribution over the values of the magnetic spin quantum number $m_s = +1/2$ and $m_s = -1/2$ provides yet another period/group equation for chemical elements:

$$\bar{I}_{ms} = -\frac{z+a}{2z}\log_2\frac{z+a}{2z} - \frac{z-a}{2z}\log_2\frac{z-a}{2z} \approx 1 - \frac{k}{(z_0+a+b)^2} \tag{2.17}$$

Here, b and the period constant z_o are those from the previous paragraph. The group constant $k = a^2/(2\ln 2)$ includes the number of unpaired electrons a. The error produced by the approximation used decreases with the fourth degree of z, and even for Li, #3, is only 0.0015.

2.5 Information Theoretic Interpretation of the Pauli Exclusion Principle and Hund's First Rule [21]

The information equations for electron distributions in atoms provided a new interpretation of some of the physical principles and rules controlling the building of the atomic electronic structure. The Hund first rule, which requires maximum filling of atomic orbitals in s–, p–, d–, and f–subshells with unpaired electrons, may be interpreted as a rule demanding maximum information on atomic orbitals, I_{nlm}. This follows directly from our Eq. (2.16), which maximizes when the number of paired electrons $b = 0$.

According to Eq. (2.10), the atomic information content reaches its absolute maximum when all $z_i = 1$. This case corresponds to electron distribution over spin–orbitals, *SO*, defined by the four quantum numbers n, l, m, and m_s, and controlled by the Pauli exclusion principle.

$$I_{SO} = z \log_2 z = \max \tag{2.18}$$

In our study [20], the above ideas were generalized by information theoretic analysis of the irreducible representations of the S_N symmetry group of the particles. It was concluded that the Pauli exclusion principle is associated with a trend toward acquiring maximum information content of atoms and molecules or any system of fermions. Conversely, the bosons were shown to be always associated with minimum information content. Another study [22] has shown that light emission from atoms is associated with a decrease, while light absorption is accompanied by an increase in the atomic information content.

2.6 Predicting the Properties of Transactinide Elements # 113–120

The information indices introduced in the foregoing describe adequately both atomic electronic structure and the structure of the Periodic Table [15,17,18]. One might expect on that basis the information indices to describe the properties of chemical elements better than the atomic charge, which presents only the total number of electrons, but not their distribution. These ideas were tested by a comparative study [23] of the atomic charge and the information indices correlations with a total of 19 physico–chemical properties of elements from groups 1, 13, 16 and 17. In 18 of the series the correlations with the information indices were higher than those with the atomic number. Additional test was performed with 37 properties of the elements from these groups. Again, with the exception of only one series, the information indices provided much higher correlation than the atomic number. The standard deviation was in some cases 5–to 10–fold smaller.

The structure–property analysis described in the previous paragraph was performed in 1979–1980. For the Kananaskis 2003 Conference [24], this part of the study was updated [25] with a wider number of properties, and the most recent data, focusing on all 23 properties known for group 1 elements: five atomic, twelve thermodynamic, and six physical properties and parameters. In all twenty–three cases examined, the correlation coefficient of the models derived with the information indices was higher than the one produced by the atomic number. In the majority of cases, the standard deviation of information models was at least 2–fold to 5–fold smaller than that of their atomic number counterparts. It went down for absolute entropy in solid state from 1.33 to 0.12 kJ/mol, for covalent radius from 6.55 to 0.51 pm, for melting point from 117.9° to 8.6°, for resistivity from 3.33 to 0.12×10^{-8} om m, for boiling point from 2.24° to 0.04°, and for the enthalpy of fusion from 0.0155 to 0.0002 kJ/mol.

The models obtained with the atomic information indices provided a basis for predicting twelve properties of unsynthesized elements from the upper end of the Periodic Table [26]. We have taken into account the relativistic effects, and additionally adjusted the information models to match better periodicity trends in groups and periods, in line with the Mendeleev tradition. It is not yet possible to verify these predictions with experimental values. However, the continuing progress in shifting up the upper end of the Periodic Table, culminating with the recent syntheses of several atoms of elements 114 and 116, has brought us closer to this goal.

2.7 Information Content of Atomic Nucleus

In describing atomic nucleus we proceeded similarly with the distribution of protons and neutrons into nuclear shells, subshells, etc., as well as according to the nuclear quantum numbers [27,28]. We briefly mention here only the information index on the proton–neutron composition of atomic nuclei, $I_{p,n}$, which exhibits interesting properties [29]:

$$I_{p,n} = A\log_2 A - z\log_2 z - n\log_2 n \tag{2.19}$$

It follows from Eq. (2.19) that the symmetric nuclei having equal number of protons and neutrons, $p = z = n$, have maximum information on the proton–neutron composition. The total information on proton–neutron composition of these nuclei is equal to mass number A, whereas the mean information is exactly 1 bit. Such are the symmetric stable nuclei of the elements from the beginning of the Periodic systems: ^{2}H, ^{4}He, ^{6}Li, etc., up to ^{40}Ca. With the increase of the atomic number of the element, the filling of the proton energy levels in the nucleus proceeds with delay. The resulting excess of neutrons is termed *isotopic number*: $\beta = n–z = A–2z$.

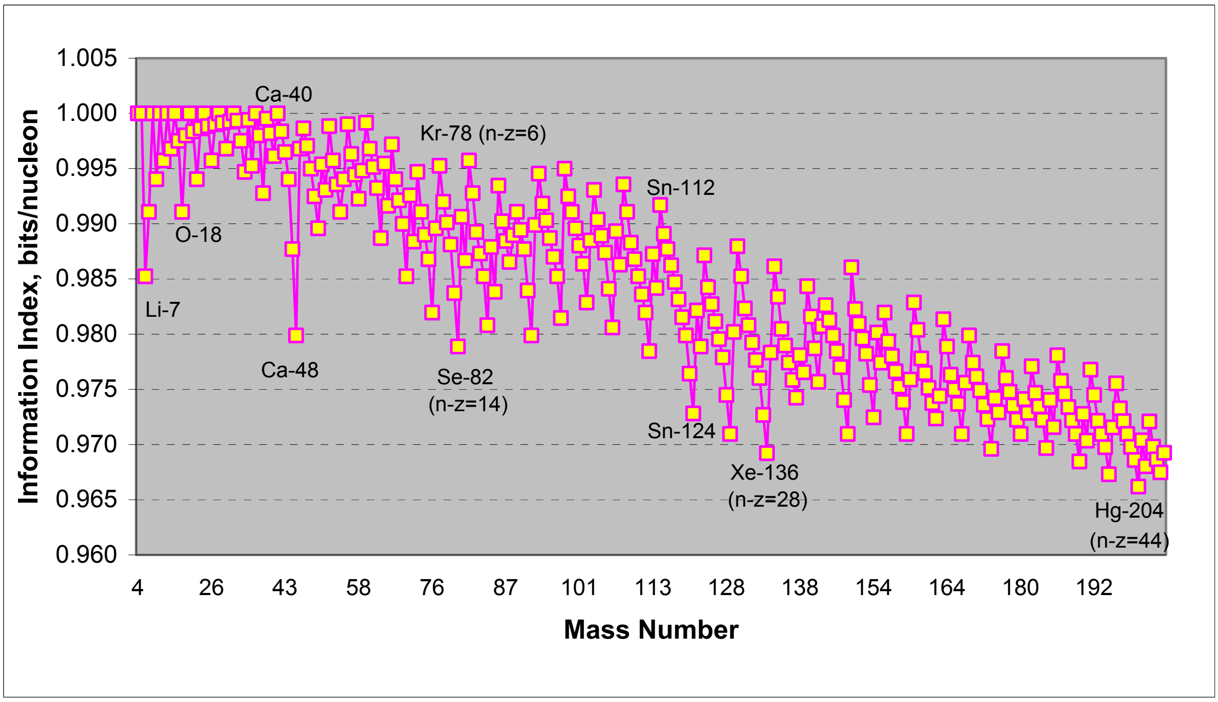


Figure 2. Average information on the proton–neutron composition of atomic nuclei.

Figure 2 illustrates the basic features of the I_{np} index for all stable nuclei up to ^{204}Hg, showing the line of nuclei with maximum mean information content of 1 bit, and the increasing deviation from this line with the increased atomic mass. The minimums in the figure correspond to the largest excess of neutrons for series of several isotopes of the same chemical element. Conversely, the maximums after ^{40}Ca refer to the least possible excess of neutrons.

The mean I_{np} index per nucleon was approximated with a sufficient accuracy to

$$\bar{I}_{np} \approx 1 - \frac{1}{2\ln 2} \times \frac{\beta^2}{A^2} = 1 - \frac{1}{2\ln 2} \times \frac{(A-2z)^2}{A^2} \tag{2.20}$$

Equation (2.20) shows that the information index on the proton–neutron composition of atomic nuclei increases with the increase in the mass number A and atomic number z, and it decreases with the increase in the isotopic number and the number of neutrons. These trends are however reversed for a small number of light nuclei having negative isotopic number.

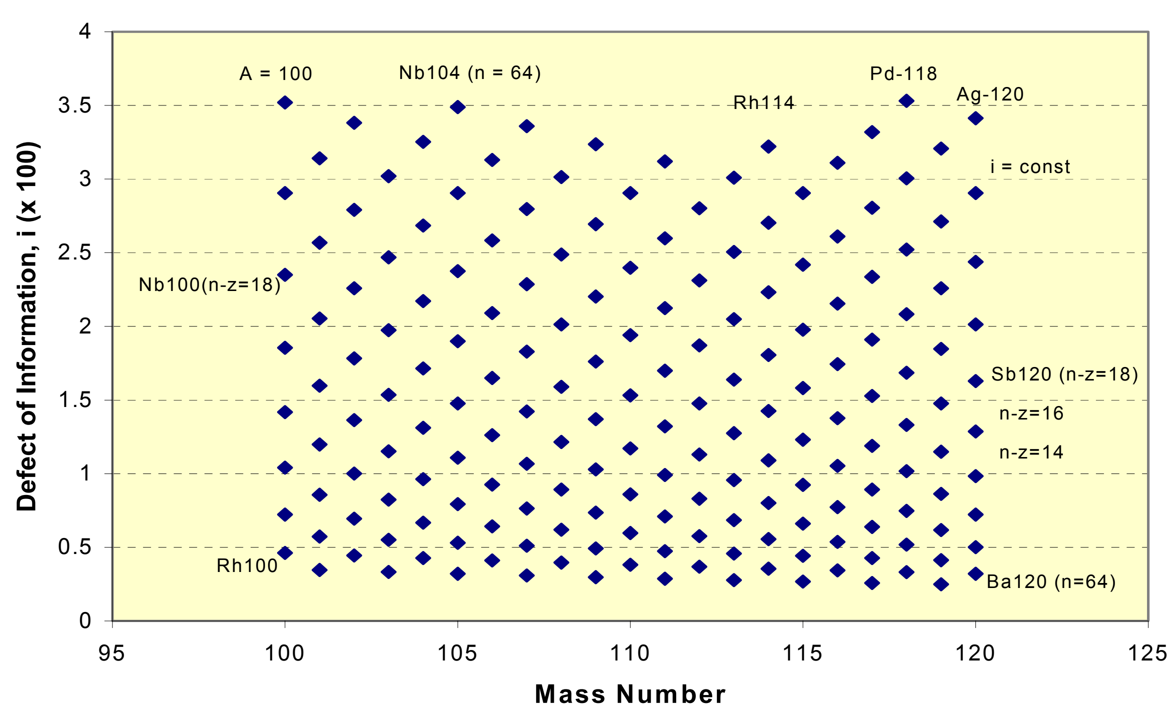


Figure 3. The defect of information on proton–neutron composition of atomic nuclei as a new Atomic Nuclei Chart, classifying the nuclides into five classes: isotopes (z = const), isobars (A = const), isotones (n = const), isodifferents ($\beta = n–z$ = const), and isodefect (i = const).

2.8 The Concept for "Defect" of Information

The total information on the proton–neutron composition of a nucleus, I_{np}, expressed in bits according to Eq (2.19), is very close to the mass number A. The difference $\Delta I_{n\mathrm{l}}^{*}$ between the two quantities was introduced in our studies [29] as "defect" of information by analogy with the defect of mass upon atomic nuclei formation. Since this deviation results in a decrease of the binding energy, we have conjectured that the defect of information can be regarded as a negative component of the binding energy, E_b. In fact, we found that $\Delta I_{n\mathrm{l}}^{*}$ coincides (with a transition coefficient k =

25.1 MeV/bit) with the parameter of the relative symmetry of the nucleus, δ, in Weizsäcker's equation for nuclear binding energy. Equations relating I_{np} and $\Delta I_{n\mathrm{l}}^{*}$ with the total nuclear binding energy were also derived for series of nuclides with a constant isotopic number β.

The concept for a defect of information also resulted in a new systematic of nuclides [28]. Besides the classes of isotopes (z = const), isobars (A = const), isotones (n = const), and isodifferents (β = const), a new class was formed for nuclides having the same mean defect of information *i*. Figure 3 shows a segment of the new Chart in defect of information/mass number coordinates (*i*/*A* diagram), which allows all five classes of nuclides to be presented in a single 2D–plot. Besides *isobars* along the vertical direction, and *isodefect* along the horizontal direction, one can also find in Figure 3 *isodifferent series* descending gradually from left to right, e.g. the β = 18 series, descending from ^{100}Nb to ^{120}Sb. The *isotone series* descend steeply in the same left to right direction, as shown with the series from ^{120}Ba to ^{105}Nb. The *isotope series* ascend steeply from left to right; the example selected is the Rh series from ^{100}Rh to ^{114}Rh. The direction and magnitude of change for these three series are described by derived information equations.

2.9 Information Equations Derived for the Four Main Classes of Nuclides [30]

	Equations	Constants
I. Isobars (A=const)		
1	$\bar{I}_{np}(isobars) = 1 - c\beta^2 = 1 - c(A-2z)^2$	$c = (2\ln 2A^2)^{-1}$
2	$\Delta I_{np}(isobars) = a - bz$	$a = \frac{2(A+1)}{A\ln 2},\ b = \frac{4}{A\ln 2}$
3	$\Delta I_{np}{}^{*}(isobars) = cA(A-2z)^2$	
II. Isodifferents (β = n–z = const)		
4	$\bar{I}_{np}(isodifferents) = 1 - \frac{d}{A^2}$	$d = \frac{\beta^2}{2\ln 2}$
5	$\Delta I_{np}(isodifferents) = 2\left[1 + \frac{d}{A(A-2)}\right]$	
6	$\Delta I_{np}{}^{*}(isodifferents) = \frac{d}{A}$	
III. Isotopes (z = const)		
7	$\bar{I}_{np}(isotopes) = 1 - \frac{1}{2\ln 2}(1 - \frac{e}{A})$	$e = 2z$
8	$\Delta I_{np}(isotopes) = f + \frac{g}{A(A-1)}$	$f = 0.2786, g = \frac{4z^2}{2\ln 2}$
9	$\Delta I_{np}{}^{*}(isotopes) = \frac{A}{2\ln 2}(1 - \frac{e}{A})^2$	

	Equations	Constants
	IV. Isotones (n = const)	
10	$\bar{I}_{np}(isotones) = 1 - \frac{1}{2\ln 2}(\frac{e'}{A} - 1)^2$	$e' = 2n$
11	$\Delta I_{np}(isotones) = f + \frac{g'}{A(A-1)}$	$g' = \frac{4n^2}{2\ln 2}$
132	$\Delta I_{np}*(isotones) = \frac{A}{2\ln 2}(\frac{e'}{A} - 1)^2$	

2.10 Predictions of Nuclear Binding Energies of Nuclides of Elements # 101–108

The correlations found between the information on the proton–neutron composition or the defect of this information, and nuclear binding energy, along with the established similarity in the information equations describing nuclides and chemical elements, prompted us to search for direct relationships between nuclear and electronic energies. Equations were derived that relate fairly well the ground state electronic energy E_e of chemical elements to the nuclear binding energy $E_b(id)$ of isodifferent series of nuclides [31].

Table 2. Comparison of the predicted and measured values of the nuclear binding energies of some nuclides of chemical elements # 101 to 108 [25]

β	*z*	*A*	E_b(pred.)	E_b(exper.)	β	*z*	*A*	E_b(pred.)	E_b(exper.)
49	103	255	1889.235	1887.465	53	106	265	1945.166	1943.199
	104	257	1899.300	1896.954		107	267	1954.800	1952.342
	105	259	1909.259	1906.121	54	102	258	1913.639	1911.128
	106	261	1919.112	1915.447		103	260	1924.599	1919.621
	107	263	1928.862	1924.336		104	262	1935.477	1930.934
50	104	258	1903.592	1904.563		105	264	1946.285	1939.257
	105	260	1912.916	1912.603		106	266	1957.026	1950.468
	106	262	1922.085	1923.259	55	102	259	1916.513	1916.569
	107	264	1931.101	1930.932		103	261	1926.384	1926.418
	108	266	1939.969	1941.345		104	263	1936.086	1936.563
51	104	259	1912.138	1910.716		105	265	1945.581	1946.224
	105	261	1921.943	1920.042		106	267	1954.913	not synth
	106	263	1931.617	1929.620	56	101	258	1910.768	1911.701
	107	265	1941.163	1938.568		102	260	1920.146	1923.139
	108	267	1950.582	1947.803		103	262	1929.240	1931.927
52	103	258	1911.461	1906.916		104	264	1938.056	1943.295
	104	260	1922.186	1918.037		105	266	1946.616	unknown
	105	262	1932.856	1926.206	57	101	259	1918.088	1917.837
	106	264	1943.456	1937.123		102	261	1928.501	1928.317
	107	266	1953.973	1944.952		103	263	1938.762	1938.413
53	103	259	1915.474	1913.955		104	265	1948.874	unknown
	104	261	1925.505	1923.949		105	267	1958.840	unknown
	105	263	1935.472	1933.417					

We were thus encouraged to use our models in predicting in the year 1980 the nuclear binding energies of unsynthesized nuclides of elements from the upper end of the Periodic Table [32]. Nine isodifferent series with β = 49 to 57, atomic numbers within the 87–103 range, and mass numbers between 224 and 257, have been used to obtain predictive models. The average standard deviation of the models was 0.13%. Forty–five extrapolations from these models were made for nuclides of elements # 101–108. Updating these results for the Kananskis Conference [24] in 2003, we found that forty–four of these forty–five nuclides have already been synthesized and the nuclear biding energies of 41 of them have been measured. The excellent agreement of the predicted and measured values is shown in Table 2 [25]. The standard deviation is only 1.9 MeV for a binding energy range of 1890–1950 MeV, i.e. the relative standard deviation is approximately 0.1%.

3 INFORMATION THEORY IN DESCRIBING MOLECULES

3.1 Prehistory

The first molecular information descriptor (index) has been introduced in 1953 by Dancoff and Quastler [33], as “information on the kind of atoms in a molecule”. The elemental composition distribution incorporates subsets of atoms of the same chemical element. The entropy of this distribution called also *information on chemical composition*, I_{cc}, calculated by eq. (2.5) is a measure of the compound *compositional diversity*. One may trace, for example, the increase in diversity within the series CH_4, CH_3F, CH_2FCl, CHFClBr to be associated with the increase in I_{cc} (3.61, 6.85, 9,61, 11.61 bits, respectively). Dancoff and Quastler have used this information index to assess the total information content of various organisms, from cell to man. A similar approach was used to characterize the composition of complex molecules like protein, proceeding from the distribution of their aminoacid residues [34].

Rashevsky proposed in 1955 another measure termed *topological information content*, proceeding from the distribution of atoms into subsets according to their equivalence in both chemical nature and symmetry, the latter being defined by the same first, second, etc. atomic neighborhood [35]. Trucco [36] refined the definition of topological equivalence of atoms in terms of the orbits of the automorphism group of the molecular graph. This type of molecular information content was later termed *orbit’s information index*, I_{orb} [37]. Mowshowitz [38] introduced in 1968 a second information index for graphs termed *chromatic information content*, I_{chr}. It is based on the graph coloring by the minimum number of colors, so that no two adjacent vertices are of the same color.

3.2 Some Measures of Molecular Information Content

The studies mentioned in the preceding subsection cover almost all that has been done in this area before the start of my work on information theoretic characterization of atoms and molecules in the year 1970. The first information index we introduced was based on grouping the atoms in a molecule into equivalence classes determined by the point group of symmetry to which the molecule belongs [39]. This *molecular symmetry index*, I_{sym}, complements the orbit's information index, in accounting for specific molecular geometry and conformations. Linear relationships were obtained between I_{sym} and thermodynamic entropy for several homologous series of organic compounds. The constant information/entropy ratio indicates that each homologous series is characterized by a constant degree of organization of atoms. The family of information indices based on the equivalence of atoms by chemical nature, automorphism symmetry and point group symmetry included later the *centric information index*, I_{centr} [40], constructed from the distribution of atoms into radial layers of increasing distance to the center of molecular graph (*vide infra*).

Besides equivalence of atoms or graph vertices, information indices can be defined from the equivalence of bond (or graph edges) and even more complex substructures. Such an idea was mentioned in the second paper of Trucco [36], without further details. Such an information index based on the automorphism group of graph edges has not been studied. The next level index, based on the equivalence of the two–edge subgraphs, has been included as a component of the first Bertz complexity index [41]. Two other different types of graph edges equivalence were also proposed [37,42]. The first one is based on the value of the partial connectivity index χ_{ij} of the edge {ij}to introduce an information analog $I\chi$ of the Randić connectivity index χ_R [43]. The number of cycles to which a vertex or an edge belongs (vertex and edge cyclic degree) was used to define two kinds of *information cycle indices* of the graph [37,42].

All information indices discussed in the foregoing are based on some equivalence relation between molecular substructures (atoms, bonds, two–bonds fragments, etc.) or their corresponding molecular graph counterparts (vertices, edges, etc.). One may, as we did in 1977 [6], to term this type of information indices, as "equivalence–based", and denote them by $^{E}I(\alpha)$, α standing for the specific equivalence relation. This type of indices is not a very good basis for structure–property or structure–activity relationships, because it is highly degenerate. To improve the applicability of information theoretic indices we looked for ways to reformulate them. Such second avenue was offered in our study with Nenad on molecular branching [6]. The alternative type of information indices was termed "magnitude–based", and denoted by $^{M}I_{\mathrm{D}}$ (originally, the notation $I_{\mathrm{D}}^{\mathrm{W}}$ was used). The equivalence–based and the magnitude–based information indices were presented later in my monograph [42] as resulted from a generalized finite probability scheme describing a system of N elements having the overall magnitude (property, weight) of M, which are distributed into classes:

Classes of elements	1	2 …	k
Number of elements	N_1	N_2 …	N_k
E–probability	$^{E}p_1$	$^{E}p_2$ …	$^{E}p_k$
Magnitude	M_1	M_2 …	M_k
M–probability	$^{M}p_1$	$^{M}p_2$ …	$^{M}p_k$

where ${}^{E}p_i = N_i / \Sigma N_i$ and ${}^{M}p_i = M_i / \Sigma M_i$.

The first index of the new type was defined for the distances in the graph. As well known, the distance d_{ij} between vertices i and j is measured by the number of edges connecting these vertices along the shortest path between them. The equivalence–based distribution of N distances into k groups having N_1, N_2, … , N_k distances, respectively, produces the standard type of Shannon's entropy by Eq. (2.5). The distance magnitude distribution is the partitioning of the Wiener number W [3] (the sum of all vertex–vertex distances), into N_1 distances of magnitude $M_1 = 1$, N_2 distances of magnitude $M_2 = 2$, … , N_k distances of magnitude $M_k = k$. The two distance information descriptors are then, respectively,

$$\begin{aligned} {}^{E}I_D &= N \log_2 N - \sum_{i=1}^{k} N_i \log_2 N_i \\ {}^{M}I_D &= M \log_2 M - \sum_{i=1}^{k} N_i M_i \log_2 M_i \end{aligned} \qquad (3.1a,b)$$

The two information indices on the distances in molecular graph have found application in characterizing molecular branching and cyclicity (*vide infra*), for isomer discrimination [37], and for structure–property and structure–activity correlations [44–50], as part of the software packages OASIS [51], CODESSA[52], and Dragon [53]. Included in these packages was also the geometric 3D–analogue of the Wiener number, WG, and its information counterpart, the I_{WG} index, devised in our Bourgas laboratory [54]. (Regrettably, the WG index, which proved to be highly significant in QSPR and QSAR models, has always been referred to (renamed as 3D–Wiener) a later work of other authors).

The extended probability scheme was applied to define several other information indices [6,42]. The *information on the vertex degree distribution*, I_{vd}, partitions the total graph adjacency $A = 2E$ (E being the total number of graph edges) into vertex degrees a_1, a_2, … , a_k, and was shown to be an important characteristics of the graph connectivity. Similar definition was proposed for the *information on the edge degree distribution*, I_{vd}, proceeding from edge adjacency matrix, as well as for the *information on the vertex distance degree distribution*, I_{dd} [42]. One of the important topological descriptors, the Hosoya index Z [4], was also reformulated as a magnitude–type information index. The Z index is defined as the sum of the non–adjacency numbers $p(G,i)$, which count the number of ways the edges of graph G can be separated by i edges. Setting $M_i = p(G,i)$ and $M = Z$ in Eq. (3.1b) one arrives at the *information index on the Hosoya graph decompositions*, I_Z [6,42].

The major application of the information theoretic indices is in constructing quantitative structure–property and structure–activity models. Such applications are discussed in Section 4.8 (*vide infra*). Another goal is to construct descriptors with low degeneracy for isomer discrimination. Since the individual indices cannot capture the entire complexity of chemical structure, we offered to construct a *topological information superindex*, *SI*, which to combine several indices representing different structural features, and expressed in the same information theory language. The six terms incorporated in the superindex represent graph adjacency (vertex degrees), distances, vertex centric distribution, vertex chromatic distribution, the Hosoya decomposition, and vertex distribution over the orbits of the automorphism group of the graph. To preserve the information each of these indices contains, they were ordered as a vector:

$$SI = \{I_{vd}, {}^{M}I_D, I_{centr}, I_{chr}, I_Z, I_{orb}\} \tag{3.2}$$

Eq. (3.2) is an open system to which other structural information descriptors might be added.

Another potential area of information theoretic indices application is for classification purposes and for complexity estimates. The latter will be discussed in the section devoted to complexity, while the next subsection illustrates a specific illustration, in which information theoretic indices were constructed for elucidating a controversial problem in theoretical organic chemistry.

3.3 The Information Index of Electron Delocalization [55,56]

The electron distribution in molecules can be described by the charge–bond order matrix **Q**. The elements q_{ij} of this matrix can be transformed into probabilities of electron distribution over atomic and molecular orbitals: $p_{ij} = q_{ij} / \Sigma q_{ij} = q_{ij} / A$ with i and j running from 1 to k. Applying Shannon's equations (2.4, 2.6a, and 2.7) one arrives at expression for the mean information content of the molecule (or molecular fragment), which measures the degree of unevenness in the distribution of electrons in the electronic state α:

$$I(\alpha) = \log_2(\frac{k}{A}) + \frac{1}{A} \times \sum_{ij} q_{ij}(\alpha) \log_2(q_{ij}(\alpha)) \tag{3.3}$$

Conversely, the evenness of this distribution may be characterized by a reciprocal index $\widetilde{I}(\alpha)$ termed *information index of delocalization* of molecule M or molecular fragment M_i in the ground or excited state α:

$$\widetilde{I}(\alpha) = I^B(0) / I^{M(or\,M_i)}(\alpha) \tag{3.4}$$

Here, I^B (0) is the reference information content of benzene in the ground state. The information index of electronic delocalization can be calculated within each LCAO–MO–CI method for an arbitrary electronic state α. It has been applied to the estimate of ring and total aromaticity in ground and excited states in a join work with the late talented Bulgarian quantum chemist Filip Fratev [55]. The information index provided the correct classifying of a large sample of conjugated compounds as aromatic, non–aromatic, and anti–aromatic within the ranges $\widetilde{I}(\alpha)$= 1 – 0.45, 0.45 – 0.30, and 0.30 – 0, respectively. A later version of the method reformulated the delocalization index proceeding from the quantum mechanical superdelocalizability indices [56]. Reference [55] was among the first papers combining information theory to quantum chemistry, an area of research that later became of considerable interest, as illustrated, for example, by the studies of Carbo [57].

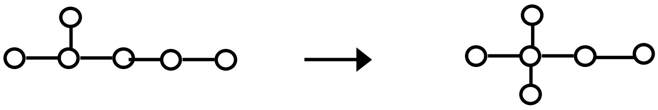
Other applications of information indices to mention deal with the classification of bioorganic compounds based on the information capacity of molecules (Zhdanov, [58]), characterization of the specificity of chemical reactions [59], optimal strategy of organic syntheses (Bertz, [41]), and others. This concludes the brief summary of the work done in chemical information theory. More details, including different interpretations of entropy and information as applied to chemical structures, may be found in my monograph [42] and a recent book chapter [60].

4 MOLECULAR TOPOLOGY

4.1 Molecular Branching [6,61–63]

Wiener [3] first analyzed some aspects of branching by fitting experimental data for several properties of alkane compounds to the diversion of his “path number” W in branched alkanes from that of the linear isomeric compound. Later, several graph–invariants were tested as “branching indices” of acyclic molecules (the Hosoya index [4], the graph largest eigenvalue [64], the two Zagreb indices [65], and the molecular connectivity index of Randić [43]), which correlate with their properties. Our work with Nenad Trinajstić on the characterization of molecular branching [6] started in the beginning of 1976. We soon realized that the Wiener index offers the excellent chance to go beyond the fitting of experimental data and to try to understand the topological basis of molecular properties. We conjectured that the Wiener number decreases with the increase in the degree of branching of isomeric molecules. Eight specific topological patterns that enhance branching were identified, proved to be associated with a decrease in the Wiener number, and to be mirrored in a number of molecular properties. They are illustrated below, along with the derived formulae for the corresponding change in the Wiener number (N and N_1 stand for the numbers of vertices in the main chain and the branch, respectively; j is the branch position in the main chain, n is the number of branches):

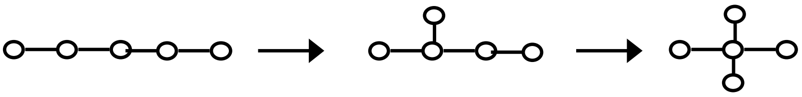
Rule 1:



(4.1)

$$\Delta W = W_1 - W_2 = (N - j - 1)(N + j - 1) > 0$$

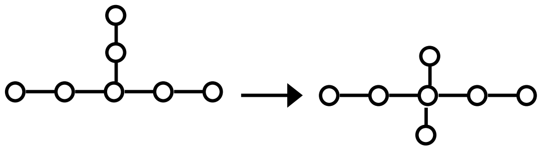
Rule 2:



(4.2)

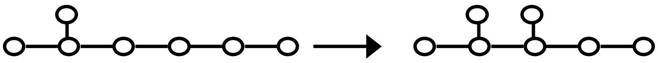
$$\Delta W = (N - j - 1)(j - 1) > 0$$

Rule 3:



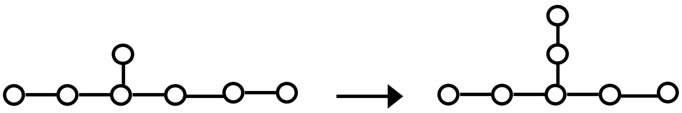
$$\Delta W = (N_1 - 1)(N - 1) > 0 \tag{4.3}$$

Rule 4:



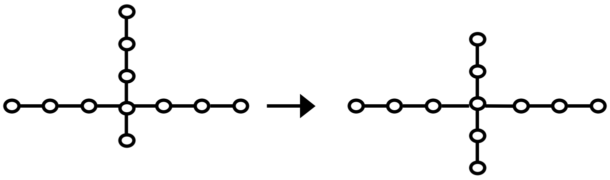
$$\Delta W = (N - j - 1)(n + j - 2) > 0 \tag{4.4}$$

Rule 5:



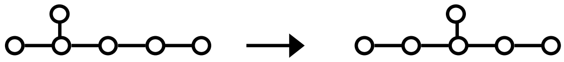
$$\Delta W = (j - 1)(N - N_1 - j - 1) > 0 \tag{4.5}$$

Rule 6:



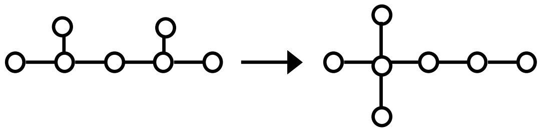
$$\Delta W = (N_1 - N_2 - 1)(N - 1) > 0 \tag{4.6}$$

Rule 7:



$$\Delta W = (j' - j)(N - j - j' + 1) > 0 \tag{4.7}$$

Rule 8:



$$\Delta W = j' - j > 0 \tag{4.8}$$

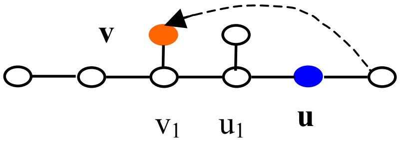
We have shown that the decrease in the Wiener index is associated with an increase in heats of formation, densities and refraction indices, and with a decrease in heats of vaporization and combustion, boiling points, molar volumes, molar refractions, and the coefficients of Antoine's vapor pressure equation. Rules 1, 3, 4, 5, and 8 control all nine properties, whereas the only exception from the Rule 2 pattern occurs for 2–methyl substituents in four properties. The only branching rule that is not followed by the properties examined is Rule 7, which requires a monotonous change when the branch is shifted from a terminal to a more central position. Our branching rules 1–8 have been later confirmed by the branching theorems of Bertz [66] based on vertex degree analysis, and by the studies of Randić [67], who used the path matrix eigenvalues.

Part of our rules overlap with later work of Ruch and Gutman [68] based on Young diagrams.

In 1995, the branching rules were generalized [63] using two theorems, which simplify the calculation of the Wiener number variation ΔW upon the graph transformations described by the branching rules [69,70]. ΔW was shown to depend [Eq. (4.9)] only on the distance numbers $d(u)$ and $d(v)$ of the vertices u and v between which a subgraph containing n_1 vertices is transferred. The second theorem simplifies further the calculation [Eq. (4.10)] using as variables the length L of the chain uv, the position i of the branches located between u *and* v, and the number of vertices in these intermediate branches, $n_{v,i}$ and $n_{u,i}$, which are located symmetrically with respect to u and v:

$$\Delta W = n_1[d(u) - d(v)] \tag{4.9}$$

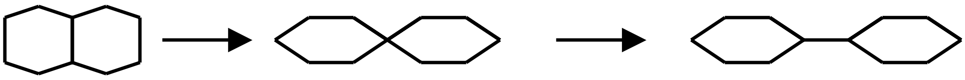
$$\Delta W = n_1 \sum_i [(L-2i)(n_{u,i} - n_{v,i})] \tag{4.10}$$



Eq. (4.10) was used to generalize the branching rules into five new rules covering more general branching patterns. Three of the new rules deal with the formation of new branches by different mechanisms, one with branch transformations related to a vertex degree redistribution, and one shows the topological identity of branch elongation to branch shifting toward a more central position. It was shown on this basis that the number of branches and the number of vertices of higher degree are considerably stronger branching factors than the branch length and branch centrality, although the role of centrality increases with the size of the system, and becomes dominant in polymeric macromolecules.

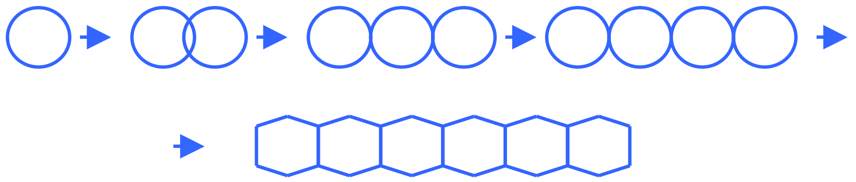
4.2 Molecular Cyclicity [71–76]

The success in describing topology of acyclic molecules prompted us to extend our studies with Mekenyan and Trinajstić on the complex topology of molecules containing atomic rings. The new concept of *molecular cyclicity* was based on a similar conjecture that all structural patterns that increase the cyclic complexity of molecules are associated with a decrease in the Wiener number. Monocyclic structures opened the series of studies, followed by three classes of unbranched polycyclic structures [71], which differ in the topology of cycles connectivity – by a common edge [72], common vertex [73], or bridge [74]

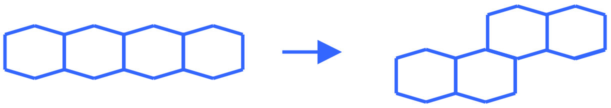


A variety of cyclicity patterns were analyzed for each of these classes, and inequalities for the associated Wiener number changes were obtained. The equations included structural factors like the number, size, and position of cycles, the number of elements (edges, vertices or bridges) they share, etc. Some of the typical molecular transformations that increase the cyclic complexity are shown below:

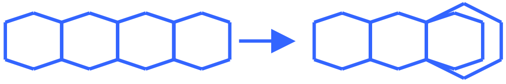
(*i*) Reduction in the cycle size for the creation of smaller cycles (1×26 → 2×14 → 3×10 → 4×8→ 6×6)



(*ii*) Transforming a linear chain of cycles into a zigzag–like one



(*iii*) Increasing the number of cycles fused to a common edge (propelerity)



The cyclicity rules for fused polycyclic compounds were confirmed later [77] proceeding from the Kirchhoff number, *Kf*, [78] as a measure of molecular cyclicity. This number is an analogue of the Wiener number, the graph distances in which are replaced by so–called "resistance distances", based on Kirchhoff's electrical network laws. Additional analysis of the problem was done recently by Randić [79]. We examined the impact of the cyclicity patterns on a number of electronic properties of aromatic benzenoids [72]. Patterns of regular variation in the HOMO–LUMO gap were found upon the molecular cyclicity rearrangements described. Thus, when the Wiener number decreases in the transformations defined by ten of our rules, the HOMO–LUMO gap increases, and for one of the rules the gap diminishes (Figure 4).

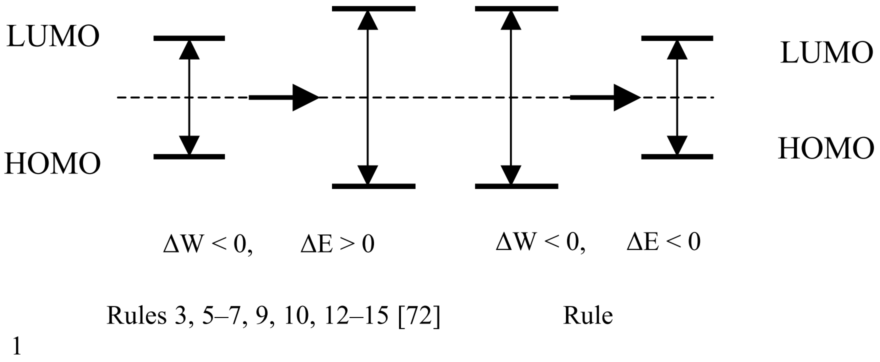
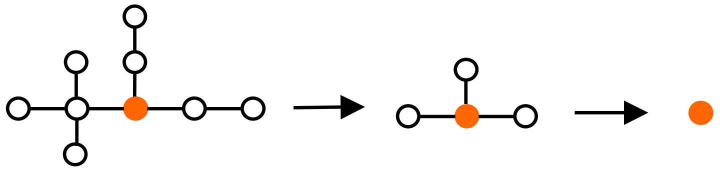


Figure 4. Increase, and respectively decrease, in the HOMO–LUMO gap of conjugated benzenoid hydrocarbons, predicted by the derived cyclicity rules, and the variation in the Wiener number.

The topology of the branched polycyclic structures was investigated last [75,76]. This class is characterized by a complex interplay between branching and cyclicity. The studies comprised the patterns of branching at a constant cyclicity, the variations in cyclicity at a constant branching, and the interconversion of cycles and branches. The rules derived were tested by eight physicochemical properties of cycloalkanes.

4.3 Molecular Centricity (Centrality)

The definition of a center of a molecule is necessarily associated with that of a graph. The first definition of a graph center given by Jordan in 1869 is limited to acyclic graphs (trees). It is based on a procedure called "pruning of a tree", which eliminates in successive steps all terminal vertices until a single central vertex or a single central edge (a bicenter) is obtained.



The general definition for a center of any graph has been given by Harary [80]. It states that the center is the vertex with the lowest *eccentricity* $e(i)$, eccentricity being defined as the longest distance from a given vertex to any other vertex in the graph. In the 1980s, attempts were undertaken jointly with Balaban, Mekenyan, and Randić [81–84], to find a more rigorous graph center definition. The definition of Harary was regarded insufficient, because it frequently classifies as central several graph vertices that are not equivalent (belong to different orbits of the automorphism group of the graph). A solution that would order centrically the graph orbits would certainly be in advantage. In the search for such an approach, we proposed a hierarchical definition, which includes several criteria [81]. Harary's definition was regarded as the first such criterion. The second criterion reduces the set of central vertices to those, which in addition to the lowest eccentricity have also the smallest sum of the distances to all the remaining vertices in the graph. This sum is termed *vertex distance*, $d(i)$, or *distance degree*, or *distasum*. When several vertices have the same minimal eccentricity and the same minimal vertex distance, then the third criterion, requiring a minimal occurrence of the largest distance, $n(i,j;\max)$, is used:

$$\text{Criterion 1:} \quad e(i) = \max(d(i,j)) = \min \tag{4.11}$$

$$\text{Criterion 2:} \quad d(i) = \sum\nolimits_j d(i,j) = \min \tag{4.12}$$

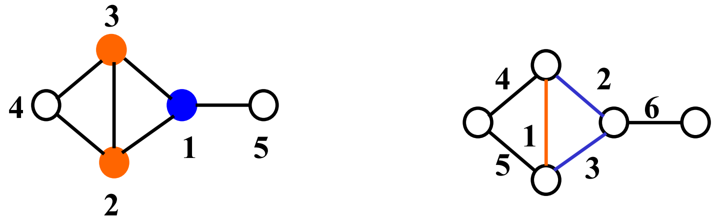
$$\text{Criterion 3:} \quad n(i,j;\max) = \min \tag{4.13}$$

Although a considerable improvement, the solution based on the hierarchical criteria 1–3 still does not always eliminate all nonequivalent vertices from the set of central vertices. Adding three more criteria, similar to (4.11)–(4.13) but based on paths instead of distances [82], also failed to solve the problem. The "ultimate" solution found [83] was based on the Iterative Vertex–Edge Centricity (IVEC) concept:

Central are those vertices that are incident to the most central edges and, vice versa, central are those edges that are adjacent to the most central vertices.

This seemingly circular definition is based on the IVEC algorithm, which resolves effectively the problem for centric ordering of the vertex and edge orbits of the automorphism group of the graph. It starts with ordering centrically all vertices and all edges, according to criteria (4.11–4.13), and ranks them numerically, beginning with the most central ones. Then, the sum of ranks of all edges incident to each vertex are calculated and used to refine the initial vertex ranks. The same is then repeated for each edge proceeding from the new ranks of their incident vertices. The steps are repeated until the same ranks appear in two consecutive steps, with one or maximum two iterations resolving the problem.

Figure 5 below illustrates the algorithm. It shows first the centric ordering of vertices and edges after criteria (4.11)–(4.13) are applied, then the centric ranks assigned to the classes of equivalent vertices and edges and, finally, the calculation of the sum of ranks in the first iteration, which determines vertices 2 and 3 as a bicenter, and edge 1 as a single central edge:



Criteria 1–3: Vertices (1,2,3), (4), (5) Edges: (1,2,3), (4,5), (6)
Ranks: 1 2 3 1 2 3
Iteration 1: v.1(2,3,6) R(1)=1+1+3 = 5; v.2 (1,3,5) R(2)=1+1+2 = 4 < 5 ⇒ Vertices 2 and 3 are centers
Centric ordering: (2,3), (1), (4), (5) (1), (2,3), (4,5), (6)

Figure 5. Illustration of the hierarchical algorithm IVEC for centric ordering of the classes of topologically equivalent of vertices and edges and rigorous defining the graph center.

We used the centric ordering of graph vertices and edges in layers equidistant from the center to construct centric topological indices [42,81]. The number of vertices in each layer forms an ordered sequence, beginning with the center, and ending in the outmost layer. For the example shown above, the vertex centric ordering obtained after applying criteria 1–3 is {3,1,1} and that of edges is {3,2,1}. This centric vertex and edge distribution is converted either into a Shannon–type *centric information index*, $^{V}I_C$ and $^{E}I_C$, or into a *radial centric index*, $^{V}C_r$ and $^{E}C_r$, respectively, the latter calculated as the sum of the corresponding squared layer cardinalities. This approach generalized a previous study by Balaban [85], which produced centric indices for acyclic graphs by applying the Jordan pruning procedure. This paper resulted from my long-term cooperation and friendship with Sandy Balaban, then a Professor at the Bucharest Polytechnics in Romania and, later, Vice-President of the Romanian Academy of Sciences. This fruitful cooperation, which started in 1979, and included many visits to Bucharest, resulted so far in 24 joint papers. It was a part of the very active "Balkan mathematical chemistry triangle", *i.e.*, Zagreb–Bucharest–Bourgas.

The centric indices contain information that is not available in other topological indices, as proved in the study of indices intercorrelation [86]. This makes the centric indices a necessary component of the regression models in QSPR and QSAR. The centric ordering of graph vertices was studied later in detail by Dobrinin *et al.* [87,88]. The ideas of centric ordering of atoms in molecules have also found application for a centric nomenclature of benzenoid hydrocarbons [89], and in developing the principles of a general centric nomenclature of chemical compounds [90]. Later work extended the centricity analysis to molecules containing multiple bonds [91], and to determining the centric ring in polycyclic molecules [92]. A recent area of application is the centrality of nodes in biological and ecological dynamic networks.

4.4. Molecular Complexity

4.4.1 Early studies

The first attempts to evaluate complexity of molecules and their graphs began 50 years ago [33], and have been based on Shannon's information theory. The information on the kind of atoms in a molecule introduced by Dancoff and Quastler (See Section 3.1) measures compositional diversity, which is one of the major components of molecular complexity. The pioneering work of Rashevsky [35] added a structural component in the assessment of molecular complexity by constructing his topological information index. Mowshowitz [12] translated these ideas into the precise language of graph theory performing an extensive information theoretic analysis of complexity of graphs. Minoli [93] was the first to propose a method for evaluating complexity of graphs without resorting to information theory as a final arbiter. His complexity measure (termed *combinatorial* complexity) is a combination of several graph invariants: the number of graph vertices V, edges E, and paths P:

$$MI = \frac{VE}{V + E} \times P \tag{4.14}$$

Unfortunately, this index is highly degenerate, particularly for acyclic graphs, for which it produces the same value for all species having a given number of vertices. In order to diminish the degeneracy, I modified the Minoli index by replacing the number of paths P by the total length of all paths, L_p [94]. Another approach, directed specifically to complexity of cyclic graphs describing chemical reaction networks, started being developing since the year 1980 jointly with Temkin and Kamenski [95] (See Section 5.4 for more detail). The method was based on the count of maximal or spanning trees (acyclic subgraphs having the same number of vertices as the graph itself). Later, the same measure was independently applied to assess complexity of graphs by Mallion and coworkers [96].

A new impulse for the study of molecular complexity came from the work of Bertz, who used as complexity measure the number of two–edge subgraphs (which he called “connections”) and its information theoretic analogue, later known as the Bertz index [41]. The success of the latter resulted from the more detailed structure description provided by the two–edge subgraphs (compared to vertices and edges used earlier) and from the additional size–term added to the Shannon formula to distinguish between molecules of different size having similar distribution of the graph elements into equivalence orbits. Bertz successfully applied his index as a criterion for selecting the critical path in complex organic syntheses. He also developed the first hierarchical concept of complexity [97] presenting it in three such levels, including both topological (branching, rings, multiple bonds) and nontopological (size, symmetry, functionality, and elemental composition) structural features. The more detailed scheme concept proposed several years later by Bonchev and Polansky [98] defines first the overall molecular complexity as determined by such factors like size, topology, physical nature of atoms, specific metric, point group of symmetry, etc. Topological complexity is then presented in full detail as a multi–layer hierarchical scheme. The basic level of *connectedness* distinguishes important classes of structures described by disconnected, connected, and nonplanar graphs. The next level of *adjacency* on its turn determines directed, undirected and multigraphs as separate complexity classes. The next level includes basic topological patterns, such as linearity, bridging, branching and cyclicity. The fourth hierarchical level makes use of graph symmetry, and the upper levels include graph metric presented by distances, paths, and walks or, alternatively, sugraphs of increasing orders [99].

The work of Bonchev and Polansky [98] also discussed in detail the requirements, which a complexity measure should satisfy, a topic first put into consideration by Minoli [93]. Later discussion by Bertz and Wright [100], Rücker and Rücker [101], Randić and Plavšić [102] and Nikolić *et al*. [103] contributed considerably to the elucidation of the problem. A consensus among mathematical chemists seems to emerge around the following criteria: A complexity measure should increase with increasing connectivity, branching, cyclicity, number of weighted vertices and edges (*e.g*., heteroatoms and multiple edges) and (at comparable other complexity features) with decreasing symmetry. Recent interest in dynamic evolutionary networks in biology and ecology [104] seems to add to the list the increased *cliquishness* (closeness of subgraphs to complete graphs having the same number of vertices), the widely used *cluster coefficient* being a measure for the cliquishness of the subgraph formed by a vertex and its first neighbors. It is still debatable whether size may be regarded as a complexity factor or rather only a complicating one. The subtle difference between the two originates in the basics of complexity theory [105,106], which consider complexity to be *relational* [107]. It describes the system as a whole, and its properties as emerged from the *interactions* between its ingredients but not reducible to their properties. The early attempts to quantify complexity did not make this distinction, when proceeding from individual information theoretic or graph theoretic indices. The more mature approaches described in the next subsection focus on the overall representation of the system as a whole.

4.4.2 Overall measures of molecular complexity

The new approach toward assessing molecular complexity emerged in the second part of the 1990s, independently, and simultaneously by the present author [108–111] and Bertz [100, 112]. It proceeds from the idea of a more complete characterization of molecular structure by counting all subgraphs of molecular graph. Such a descriptor termed *subgraph count* (*SC*) has been proposed in a single sentence a decade earlier in a paper by Herndon and Bertz, devoted to molecular similarity [113]. Bertz applied the subgraph count as a criterion for selecting the critical path in organic synthesis as the path of maximum gain in complexity. My work was oriented toward transforming the *SC* index (which I initially called “complexity index *K*”) into an optimized molecular descriptor for QSPR and QSAR. For that purpose, the subgraph count was presented as a sequence of terms, counting the subgraphs of increasing size, beginning with ${}^0SC$ for the number of vertices, ${}^1SC$ for the number of edges, ${}^2SC$ for the number of two–edge subgraphs, etc.:

$$SC={}^0SC+{}^1SC+{}^2SC+\ldots+{}^ESC \tag{4.15}$$

The partial subgraph counts, ${}^eSC$, where e stands for the constant number of edges, were tested with success as individual descriptors in QSPR studies [109,110]. The quality of property models derived was increased further [111,114] by partitioning the subgraphs from a given *e*–class into several terms describing species with different topology, like paths, stars, and triangles for subgraphs having three edges. Indeed, the latter improvement of our approach was strongly influenced by the molecular connectivity series of descriptors of Kier and Hall [115, 116], which made an epoch in the area of drug design.

In my work, the subgraph count was regarded only as the first class of descriptors in the series of *overall topological indices*. The idea for characterizing molecular complexity dealing with the structure as a whole was more general, and regarded the subgraphs weighted by some structure invariant, and then summed–up to produce the corresponding overall index. The most natural invariant to be considered was the subgraph total adjacency (the sum of vertex degrees). The resulting index was initially [108–110] termed somewhat broadly *topological complexity*, *TC*, and later renamed to *overall connectivity*, *OC*:

$$OC={}^0OC+{}^1OC+{}^2OC+\ldots+{}^EOC \tag{4.16}$$

Other graph invariants were also applied to generate overall indices, such as the *overall Wiener index*, *OW* [117], based on the Wiener indices [3] of the subgraphs, and the *overall Zagreb indices*, *OM*1 and *OM*2 [118], based on the Zagreb indices *M*1 and *M*2 [65], the predecessors of molecular connectivity indices:

$$OW={}^1OW+{}^2OW+\ldots+{}^EOW \tag{4.17}$$

$$OM={}^0OM+{}^1OM+{}^2OM+\ldots+{}^EOM \tag{4.18}$$

All overall indices were shown to correlate highly with the physicochemical properties of alkanes, used as benchmark, the overall connectivity indices being #1. This successful match stems from the properties of the overall complexity measures, which were shown to vary regularly with such complexity patterns like increasing branching, cyclicity, and centrality (See examples below).

The *total walk count*, *TWC*, introduced by Rücker and Rücker [101,119], is another important measure of molecular complexity. Being not based on counting subgraphs, it may still be regarded as another kind of overall complexity descriptor, with individual contributions from walks of different length l, ${}^{l}WC$:

$$TWC = {}^{1}WC + {}^{2}WC + {}^{3}WC + \ldots + {}^{V-1}WC \qquad (4.19)$$

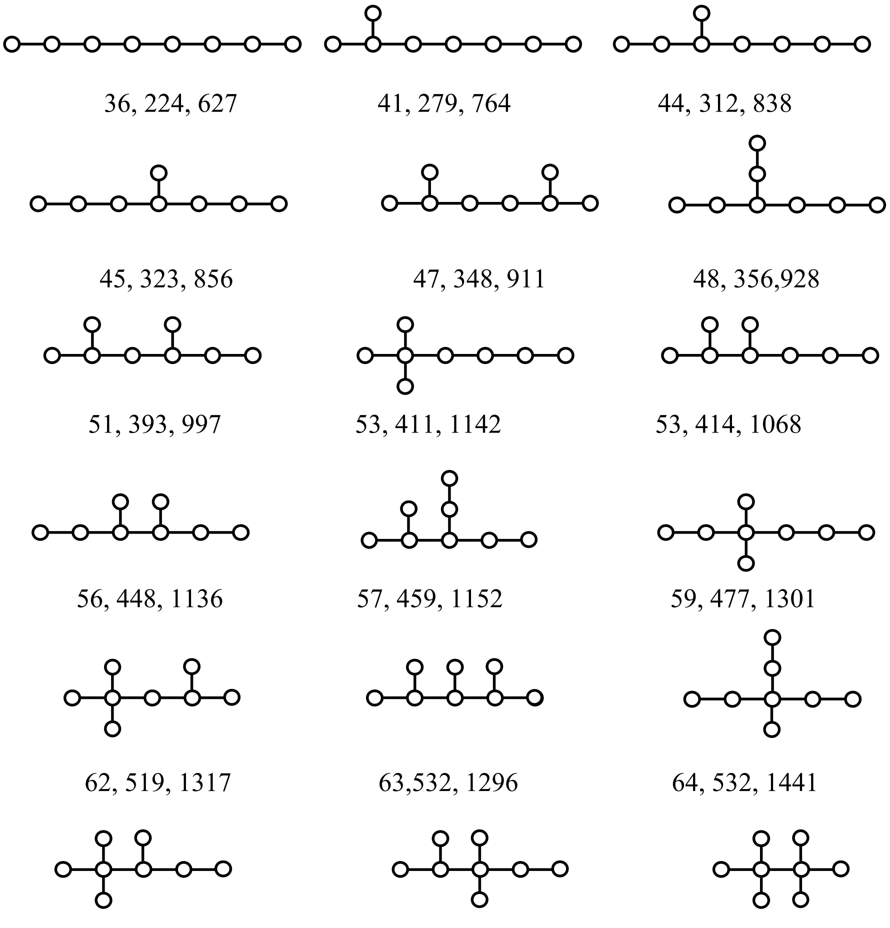

36, 224, 627
41, 279, 764
44, 312, 838
45, 323, 856
47, 348, 911
48, 356,928
51, 393, 997
53, 411, 1142
53, 414, 1068
56, 448, 1136
57, 459, 1152
59, 477, 1301
62, 519, 1317
63,532, 1296
64, 532, 1441
69, 597, 1536
71, 618, 1609
86, 798, 2047

Figure 6. Subgraph Count (*SC*), Overall Connectivity (*OC*), and Total Walk Count (*TWC*) order similarly the 18 octane isomers, according to their increasing complexity (increasing number, size, and more central position of their branches).

As seen in Figure 6, subgraph count, overall connectivity, and total walk count are very sensitive toward subtle variations in molecular topology, satisfy the criteria formulated above for a complexity measure, and assess complexity of molecules in a rather similar manner.

4.4.3 Overall measures of molecular complexity

Shown below are some formulae we derived [114] for the subgraph count and overall connectivity of some basic classes of graphs having chemical relevance (v and e stand for the number of vertices and edges in the respective subgraphs, while V and E are the total number of vertices and edges in the graph, respectively).

Path Graphs (P_V):

$$^{e}SC(P_V) = V - e \quad ; \quad SC(P_V) = \frac{V(V+1)}{2} \tag{4.20a,b}$$

$$^{e}OC(P_V) = 2[E(e+1) - e^2] \quad ; \quad OC(P_V) = \frac{V(V-1)(V+4)}{3} \tag{4.21a,b}$$

Cyclic Graphs (C_V):

$$^{e}SC(C_V) = V - e \quad ; \quad SC(C_V) = V^2 + 1 \tag{4.22a,b}$$

$$^{e}OC(C_V) = 2V(e+1) \quad ; \quad OC(C_V) = V(V^2 + V + 2) \tag{4.23a,b}$$

Star Graphs (S_V):

$$^{e}SC(S_V) = \binom{E}{e} \quad ; \quad SC(S_V) = 2^E + E \tag{4.24a,b}$$

$$^{e}OC(S_V) = (E+e)\binom{E}{e} \quad ; \quad OC(S_V) = 2^E + \sum_{e=1}^{E}(E+e)\binom{E}{e} \tag{4.25a,b}$$

The equations given in the foregoing define *global* complexity measures $X(G)$. For comparative purposes one may also use *relative* measures, $X_r(G)$ or $^kX_r(G)$, obtained by dividing the global index by the value that index has for the complete graph K_V having the same number of vertices V. Bertz and Herndon [113] have first used such an approach to compare similarity of a molecule to the most complex molecule of the same size. We defined [120] similarly the kth order relative complexity index, $^kX_r(G)$, by dividing the k^{th} order index $^kX(G)$ by the k^{th} order of X in the respective complete graph:

$$X_r(G) = \frac{X(G)}{X(K_V)} ; \; ^kX_r(G) = \frac{^kX(G)}{^kX(K_V)} \tag{4.26a,b}$$

Noteworthy, the approximate complexity measure of networks called *connectance* [121,122] or (more correctly) connectedness (*Conn*) [123,124] appears naturally within our scheme as a relative edge complexity E_r (or relative adjacency A_r):

$$E_r = \frac{E}{E(K_V)} = \frac{2E}{V(V-1)} = Conn = \frac{A}{A(K_V)} = A_r \tag{4.27}$$

Average complexity per vertex and per edge, which can be of use in large networks, were also defined [120]:

$$X(vertex) = X(G)/V \; ; \; ^kX(vertex) = {}^kX(G)/V \tag{4.28a,b}$$

$$X(edge) = X(G)/E\,;\ {}^{k}X(edge) = {}^{k}X(G)/E \tag{4.28c,d}$$

Developed initially as measures of molecular complexity, overall complexity indices may find application in assessing complexity of biological networks [123–127] for a broad range of comparative and evolutionary studies. Due to the large size of these networks, containing hundreds or thousands of nodes, it suffices to use only the first three terms in Eqs. (4.15), (4.16), and (4.19). It is also preferable to use relative or normalized measures, as defined by Eq. (4.26), rather than global complexity measures. For this purpose, we derived equations for the subgraph count and overall connectivity terms of subgraphs in the complete graph having up to three–edges [127]:

$${}^{2}SC(K_V) = E \times (a_i - 1) = \frac{1}{2}V(V-1)(V-2) \tag{4.29}$$

$${}^{3}SC(K_V) = \frac{1}{6}V(V-1)(V-2)(4V-11) \tag{4.30}$$

$${}^{3}SC(K_V, triangle) = \frac{1}{6}V(V-1)(V-2) \tag{4.31}$$

$${}^{3}SC(K_V, linear) = \frac{1}{2}V(V-1)(V-2)(V-3) \tag{4.32}$$

$${}^{3}SC(K_V, star) = \frac{1}{6}V(V-1)(V-2)(V-3) \tag{4.33}$$

$${}^{1}OC(K_V) = V(V-1)^2 \tag{4.34}$$

$${}^{2}OC(K_V) = \frac{3}{2}V(V-1)^2(V-2) \tag{4.35}$$

$${}^{3}OC(K_V) = \frac{1}{6}V(V-1)^2(V-2)(16V-45) \tag{4.36}$$

$${}^{3}OC(K_V, triangle) = \frac{1}{2}V(V-1)^2(V-2) \tag{4.37}$$

$${}^{3}OC(K_V, linear) = 2V(V-1)^2(V-2)(V-3) \tag{4.38}$$

$${}^{3}OC(K_V, star) = \frac{2}{3}V(V-1)^2(V-2)(V-3) \tag{4.39}$$

Other measures of network complexity have also been introduced and tested. The information on the vertex degree distribution, I_{vd},

$$I_{vd} = \sum_{i=1}^{V} a_i \log_2 a_i \tag{4.40}$$

was also shown to match closely *SC*, *OC*, *and TWC* complexity patterns [123,125].

A new, combined complexity measure was proposed very recently. Networks with high complexity are characterized by both high vertex–vertex connectedness and small vertex–vertex separation (the "small–world" concept of Watts and Strogatz [122]). Therefore, it seems logical to use jointly the two quantities in characterizing network complexity. The ratio $A/D = <a_i>/<d_i>$ of the total adjacency and the total distance of the graph or, equivalently, the ratio of the average vertex degree $<a_i>$ and the average distance degree $<d_i>$, may be regarded as a logical approach to such a complexity measure [126]. At a constant number of vertices, the *A/D* index has a minimum value in path graphs, P_V, which are characterized by low connectivity and long distances. In contrast, the *A/D* ratio has a maximum value in the complete graphs, K_V, which are maximally connected and all of their vertices have only a unit distance separation. The classes of star graphs, S_V, and monocyclic graphs, C_V, are of intermediate complexity and their *A/D* indices are between these two extremes:

$$A/D(P_V) = \frac{2(V-1)}{V(V-1)(V+1)/3} = \frac{6}{V(V+1)} \tag{4.41}$$

$$A/D(K_V) = \frac{V(V-1)}{V(V-1)} = 1 \tag{4.42}$$

$$A/D(S_V) = \frac{2(V-1)}{2(V-1)^2} = \frac{1}{V-1} \tag{4.43}$$

$$A/D(C_V, odd) = \frac{2V}{2V(V^2-1)/8} = \frac{8}{(V^2-1)} \tag{4.44a}$$

$$A/D(C_V, even) = \frac{2V}{V^3/4} = \frac{8}{V^2} \tag{4.44b}$$

It follows from Eq. (4.42) that all graphs have their *A/D* values within the 0 to 1 range. Like all normalized complexity indices this index decreases rapidly with the graph size for path graphs, monocyclic graphs, and other weakly connected graphs, the distance in which dominates strongly over adjacency. Another version of the A/D index can be used when more subtle topological features are to be characterized. The ratio $b_i = a_i / d_i$ of the vertex degree a_i and its distance degree d_i is a local invariant with interesting centric properties. It is ≤ 1, the equality occurring for the central vertex in the star graphs, as well as for every vertex in the complete graph. The sum over the b_i values of all graph vertices behaves similarly to the *A/D* ratio, with less degeneracy and more sensitivity to local topology. We defined this sum as a new complexity index *B*, more details about which are given elsewhere [126]:

$$B = \sum_{i=1}^{V} \frac{a_i}{d_i} \tag{4.45}$$

4.4.4 Complexity of directed networks

In contrast to the extensive analysis of complexity of undirected graphs, there were no studies dealing with complexity of directed graphs. The problem was of importance for assessing complexity of biological and ecological networks. It was shown in our recent publications [124,126] that substructure count, overall connectivity, walk count and information on vertex degree distribution can again be used for assessing complexity of directed graphs, after taking into account the partitioning of the vertex degree into in–degree and out–degree. Another specifics of directed graphs is that there is no more a complete vertex–vertex connectedness in both directions; some vertices cannot be reached from other vertices. A useful characteristic of directed graph was introduced [124] and termed *vertex accessibility*. It accounts for the degree to which the vertices in directed graphs are mutually accessible by counting the pairs of vertices connected by a finite path. More specifically, the vertex accessibility $Acc(DG)$ of directed graph DG is defined as the ratio of the number of finite distances in the directed graph, $N_d(DG)$, and the total number of distances in the parent undirected graph $N_d(G)$:

$$Acc(DG) = \frac{N_d(DG)}{N_d(G)} \tag{4.46}$$

We used Eq. (4.46) to produce a more realistic estimate of the average degree of separation of a pair of vertices in a directed graph. Dividing the average distance in a directed graph, $<d> = D/N_d$, by vertex accessibility, one normalizes this quantity to the case of complete vertex accessibility. The *adjusted average distance* (*adjusted average path length*), $AD(DG)$:

$$AD(DG) = \frac{<d>}{Acc} = \frac{D \times V^2}{N_d{}^2} \tag{4.47}$$

thus defined, is larger than the average distance in the parent undirected graph, and can be used for comparisons of the average degree of separation in directed graphs.

Vertex accessibility was also used to define a more realistic measure of directed graph connectedness, as the simplest complexity measure. The new measure might be termed *accessible connectedness*, $AC(DG)$:

$$AC(DG) = C(G) \times Acc(DG) = C(G) \times \frac{N_d(DG)}{N_d(G)} \tag{4.48}$$

4.5 Topology of Crystals

This section reviews the series of joint studies [128–136] of the Bourgas Laboratory of Mathematical Chemistry (Bulgaria) and Dr. Hans Fritsche from Friedrich Schiller University in Jena, Germany, performed during the years 1979–1989. Dr. Fritsche was awarded for these studies with a special award for originality of research of the State Agency for Science of East Germany. The main idea of these studies is that it is topology that determines the equilibrium forms of crystals and atomic clusters. The criterion used in most studies was the minimum atom–atom separation in terms of minimum total distance of the corresponding graph, as expressed by the Wiener number [3], which was shown to predict well the most compact (the most stable) structure. Alternatively, the maximum total adjacency was used as stability criterion in a number of cases, both criteria showing similar structural patterns.

4.5.1 Crystal growth [128]

This idea was tested first with three models of crystal growth. The detailed sequences of crystal growth were constructed by adding an atom at each step, and by selecting from a number of candidate–structures the one with the minimum Wiener number. The first model reproduced a shape maximally close to the spherical shape typical for the free nucleation in vapor phase, and crystallization under zero–gravity conditions (Figure 7).

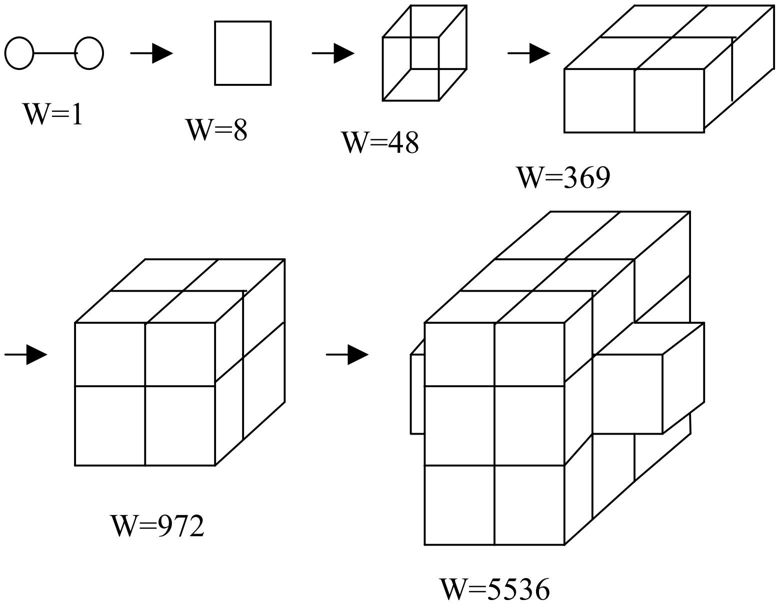


Figure 7. Selected steps of the Wiener–number–based modeling procedure of crystal growth in vapors or zero gravity conditions. The structures shown are those with the minimum value of the Wiener number at given number of atoms.

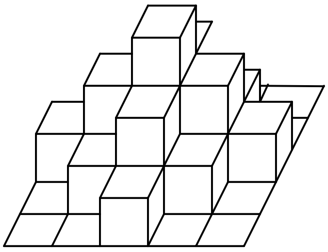


Figure 8. Three–step pyramid built as a Wiener–number–based model of crystallization on a substrate with low surface energy.

The second model reproduced nicely crystallization on a substrate with a low surface energy. The best model was built by allowing the crystal to grow on a very limited surface. The Wiener number was calculated by including also the distances to all surface lattice points in addition to those between the atoms of the growing crystal. A multi–step pyramid resulted in agreement with the periodic bond–

chain–vector crystal model and the experimental findings (Figure 8).

The third model reproduced successfully the crystallization on a substrate with a high surface energy, a result obtained by involving interaction with a considerably larger surface area. The criterion of minimum Wiener number in this case restricted the crystallization in a monolayer in agreement with the Kossel–Stransky theory and experimental observations.

4.5.2 Modeling of Crystal Vacancies and Defect Atoms [129–131]

A challenging idea was to try to predict the most probable locations of crystal vacancies [129,130] and defect atoms [131]. This goal was pursued by maximizing the difference in the Wiener number with and without vacancy (defect) as a selection criterion:

$$\Delta W = W - W_0 = \max \tag{4.49}$$

Equations were derived for a series of two– and three–dimensional models of crystal lattice with variable vacancy locations. Thus, for a simple cubic crystallite having N = 3x3x3 atoms, the variation in the Wiener number is expressed as:

$$\begin{aligned}\Delta W = \frac{1}{2}\{3N^4 + N^3[1-2(i+j+k)]\} + N^2(i^2+i+j^2-j+k^2-k) - \\ -2N(i+j+k-2) + 2(i^2+i+j^2-j+k^2-k)\end{aligned} \tag{4.50}$$

where *i*, *j*, *k* are the lattice nodes along the *x*, *y*, and *z* coordinate axes, respectively. The analysis showed that ΔW has a minimum when the vacancy is located in the lattice center and it increases when going from volume to face to edge to corner. The predicted location of vacancies on the corners of the crystallite is in agreement with thermodynamic theory and quantum chemical calculations. The same approach was used to model the diffusion of vacancies and defects through the lattice, and even for semi–quantitative estimates of the activation energy of these migrations.

4.5.3 Modeling of Atomic Clusters [132–135]

The Wiener number minimum was used again as a criterion for determining the optimum topology of small clusters [135]. Seven basic polyhedra were used in constructing small clusters: tetrahedron, square pyramid, octahedron, pentagonal pyramid, trigonal prism, centered hexagon, and pentagonal bipyramid. Adding one atom at a time over a certain crystal face and connecting this atom to all face atoms built cluster genetic lines. Two of the genetic lines produced an icosahedron, two others yielded a cubo–octahedron, and another line generated an anticubo–octahedron. Our predictions agreed with the experimental data for all cluster sizes. Thus, the topologically preferred shape for clusters with three atoms is the cyclic one, which is observed experimentally for Sc_3, Cu_3, Li_3, Ag_3, and Na_3. The ESR data for Be_4, Mg_4, and Ca_4 also confirmed our results. Other theoretical methods predict different cluster geometry for Li_4, Na_4, K_4, and Al_4. Our method predicted the pentagonal bipyramid as the most stable clusters with seven atoms in agreement with the experimental data and quantum chemical calculations for Li_7, Na_7, and K_7. The minimum of the Wiener number in the icosahedral cluster also explained the “magic” number 13, for which a maximum intensity of cluster mass spectra has been observed. Similarly, our topological approach predicted correctly the doubly magic metal superclusters $[(M_{13})_{13}]_n$, where M = ruthenium, rhodium or gold [133], as well as the stable argon clusters at the magic numbers 13, 19, 23, 26, 29, and 32 [132,134]. The Wiener number variation was also used to predict correctly the ordered structure of absorbed gases in host lattices [136].

4.6 Topology of Polymers

4.6.1 The TEMPO approach [137–141]

In the early 1980s, we started a series of studies extending the topological ideas of branching and cyclicity from small molecules to polymers. The direct approach, which uses the Wiener number as a criterion, was not applicable, due to the very large size of polymer macromolecules. We offered a modified version of the Wiener number, called “Wiener infinite”, W_∞, which has a finite value for an infinitely large polymer chain [137]. Later, the method was named TEMPO, an abbreviation for Topological Extrapolation Method for Polymers [138]. The Wiener infinite index, was defined as the limit of the Wiener number of a polymer having N atoms, normalized by dividing it by the total number of distances in the corresponding polymer graph, $N(N–1)/2$, and the total number of bonds, $E = N+C–1$, where C is the number of independent graph cycles:

$$W_\infty = \lim_{\lim \to \infty} \frac{aN^3 + bN^2 + cN + d}{[N(N-1)/2](N+C-1)} \tag{4.51}$$

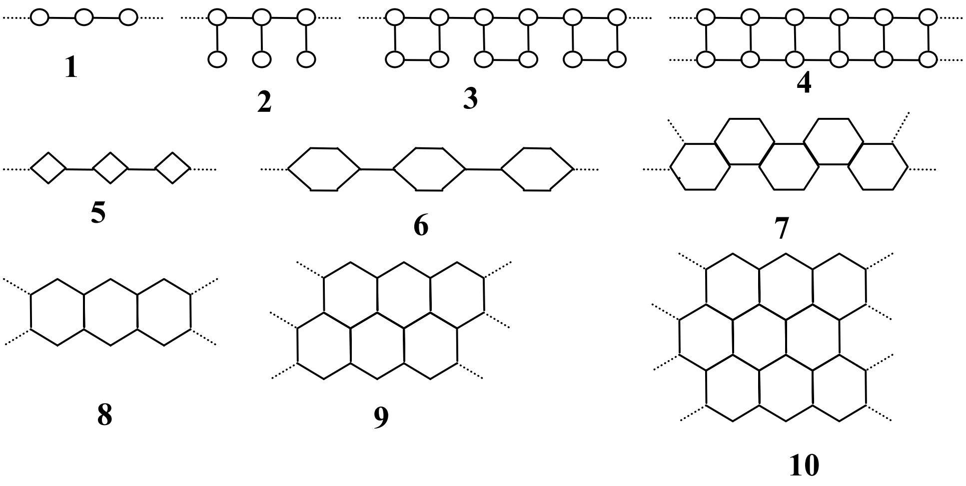


Figure 9. Nine polymer series analyzed by the TEMPO method.

The novel index has thus the meaning of average graph distance per polymer bond. The polynomial in the numerator of Eq. (4.51) is obtained from a distance matrix analysis of the initial members of the polymer–homologue series. The index preserves the patterns of the Wiener index to decrease with increasing branching and cyclicity of molecular skeleton, which is a basis for polymer structure–property relationships. A larger number of polymer structures, including acyclic, as well as cata–, peri–, and corona–condensed polycyclic structures were analyzed by the TEMPO technique. One such series is shown in Figure 9, along with an illustration of Eq. (4.51) for structure **9**:

$$W_\infty = \lim \frac{(N^3 + 8N^2 + 35N - 188)/18}{4N^3 - 11N^2 + 7N)/6} = \frac{1}{12} = 0.0833 \qquad (4.52)$$
$$\lim \to \infty$$

The Wiener infinite index correlates well with the electronic characteristics of polycyclic conjugated compounds. Thus, for a large series of 25 such compounds a linear correlation with the π–electron energy was obtained:

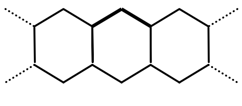
$$\begin{gathered} E_{\pi,\infty} = 1.533 - 0.851\ W_\infty \\ n = 25, r = 0.902, s = 0.01 \end{gathered} \qquad (4.53)$$

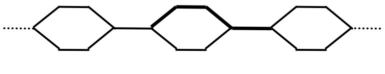
The HOMO–LUMO gap also correlates well with W_∞, and a search was done (in collaboration with Oskar Polansky) for some modifications of polymer topology that would lead to a zero gap, as a prerequisite for a potential organic superconductivity [139]. Work on other applications of the technique developed showed that some properties of industrial polymers can be predicted with a reasonably good accuracy [140]. The method may be regarded as a viable alternative to the traditional methods of calculating polymer properties based on group–additive methods or the Padé approximants, which require a large number of empirical parameters.

Recently, a joint study with Teodor–Silviu Balaban and Alexandru Balaban [141] revisited and substantially improved the method. Eq. (4.51), which requires derivation of polynomials, was transformed into a simple equation incorporating only topological invariants of the monomer unit. These are the numbers of atoms N_1 and cycles C_1 in the monomer unit, as well as the number of bonds d (or the graph distance) between two neighboring monomer units:

$$W_\infty = \frac{d}{3(N_1 + C_1)} \qquad (4.54)$$

Examples:





$D = 2, N_1 = 4, C_1 = 1, W_\infty = 2/15$; $\qquad d = 4, N_1 = 6, C_1 = 1, W_\infty = 4/21$

4.6.2 New Types of Polymer Graphs and Derived Wiener Index Equations [142,143]

This is a research performed during the period of 1997–2001 for the Exxon Chemical Polymer

Research Center in Houston in collaboration with Armen Dekmezian and Eric Markel. We introduced several different types of polymer graphs for modeling properties of polymers obtained by different mechanisms of polymerization. The preliminary analysis of polymers with a statistical distribution of their branches by number, size and position along the macromolecule main chain has shown that counter to intuition the more random the polymer structure, the simpler the graph that adequately describes polymer properties. We have thus defined *Random Distribution Graph* (*RDG*) as a graph containing only branched and terminal vertices. The *Uniform Distribution Graph* (*UDG*) matches a favorite model in polymer theory, in which all macromolecule chains (branches, spacers separating branches, and end fragments) are of equal length. The third type of polymer graph, termed *Long–Chain–Branch Graphs* (*LCBG*), was introduced to describe the class of *LCB* polymers developed during the last ten years, characterized by long branches of equal length, and a constant branch/spacer length ratio (Figure 10).

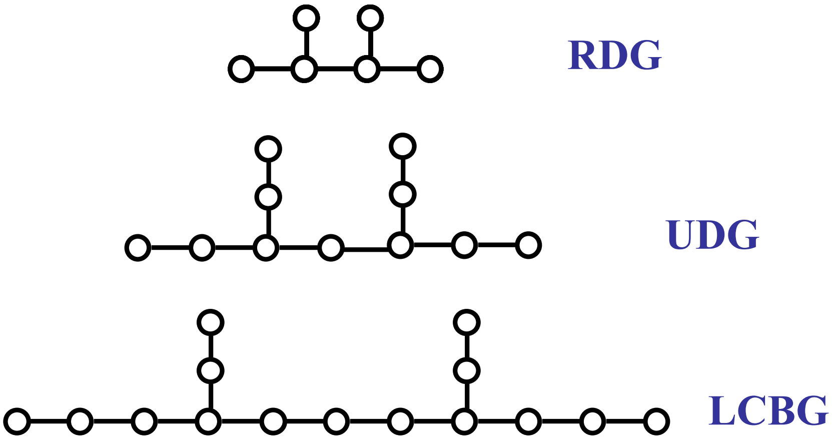


Figure 10. Illustration of the three types of polymer graphs introduced. Random distribution graphs contain branched and terminal vertices only. All branches, spacers, and terminal chains in uniform distribution graphs are of the same length. The long–chain–branch graphs have branches of equal length and a constant branch/spacer ratio.

We derived equations for the Wiener number of these classes of polymer graphs, and some of their most important subclasses. Two of these equations are shown below:

$$W(RDG) = \frac{1}{6}(f-1)^2 B(B-1)(B+7) + B(f^2-1) + 1 \tag{4.55}$$

$$W(LCBG) = \binom{n+1}{3} - \frac{1}{6} bB\{(s+b)[(s+b)(B+1)(B+2) - 6bB] - 6bs\} \tag{4.56}$$

where B is the number of branches, f is the vertex degree of the branched vertex ($f = 3$ or 4), b and s are the number of vertices in the branch and the spacer, respectively.

4.6.3 Theorems linking the Wiener number to the radius of gyration and viscosity of polymer melts and solutions

We have used the Wiener number to define the *topological radius of a* (*macro*) *molecule*, as the average distance in the respective molecular or polymer graph. More precisely, this is the total distance of the graph averaged over all N^2 distance matrix entries (the Wiener number is a sum of the $N(N-1)/2$ entries in the symmetrical off–diagonal submatrix):

$$R_{top} = \frac{2W}{N^2} \tag{4.57}$$

This definition was used to prove the following theorems:

Theorem 1: The mean–square radius of gyration R_g^2 of a macromolecule containing no atomic rings is proportional to its topological radius:

$$R_g{}^2 = \frac{b^2 R_{top}}{2} = \frac{b^2}{N^2} W \tag{4.58}$$

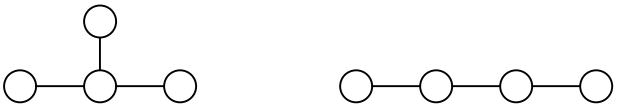
where b is the length of the covalent bond connecting two monomer units.

Theorem 2: The Zimm–Stockmayer [144] branching ratio g of a branched macromolecule containing no atomic rings is equal to the normalized Wiener index W' of the respective polymer graph:

$$g = \frac{R_g{}^2}{R_{g,lin}{}^2} = \frac{W}{W_{lin}} = W' \tag{4.59}$$

where W_{lin} and $R_{g,lin}^2$ are the Wiener number and the mean–square radius of gyration of the linear polymer having the same molecular weight with the branched macromolecule, respectively.

This is an important result, which makes it possible for the first time to measure a topological index experimentally (R_g^2 and g are measured by laser light scattering). This result sheds also light on the meaning of topology and molecular graphs. The example shown below illustrates the point:



g (3–arm star) = W / W_{lin} = (3x1+3x2) / (3x1 + 2x2 + 1x3) = 9/10 = 0.9

The simple calculation of the Wiener number of the two random distribution graphs produces immediately the well–known experimental value $g = 0.9$ of statistical 3–arm star polymers. This value does not depend on the polymer molecular weight or the branch length and position; it depends solely on the star–like topology. The same value of 0.9 is obtained in the classical theory of polymer dimensions of Zimm and Stockmayer [144] after a complicated averaging of 3D–distances over all possible configurations that the singly branched macromolecule could have. How is it possible to replace this complex procedure by the simple calculation shown above? My answer to this question is: because **topology may be regarded as average metric**.

We presented a second, more direct derivation of Eq. (4.58) within the framework of the viscoelastic theory of polymers of Rouse and Zimm [145]. The two basic matrices used in this theory, the Rouse matrix and the Zimm matrix, are in direct correspondence to the adjacency matrix and the Kirchhoff (or Laplacian) matrix in graph theory. We made use of the formula relating the mean–square radius of gyration with the eigenvalues λ_i of the Zimm matrix [146]:

$$R_g{}^2 = \frac{b^2}{N}\sum_{i=1}^{N-1}\frac{1}{\lambda_i} \tag{4.60}$$

b and *N* having the same meaning as in Eq. (4.58). Combining (4.60) with the formula connecting the Wiener number of acyclic graphs to λ_i, one arrives at our Eq. (4.58). Continuing along this line we obtained a direct link between the Wiener number and the zero–shear viscosity of acyclic polymers melts:

$$\eta_0 = \frac{cb^2\xi}{6N^2}W \tag{4.61}$$

where ξ is the friction coefficient, and *c* is the number of polymer chains in a unit volume. A generalization of our formulae (4.58) and (4.61) was made to polymer containing cycles of atoms by replacing the Wiener number by the Kirchhoff number, *Kf*:

$$R_g{}^2 = \frac{b^2}{N^2}Kf \qquad ; \qquad \eta_0 = \frac{cb^2\xi}{6N^2}Kf \tag{4.62}$$

It should be mentioned that our Eqs. (4.58), (4.61), and (4.62a,b) are the first equations connecting topological indices (the Wiener number and the Kirchhoff number) directly to physico–chemical properties, tracing the way for future development of chemical graph theory applications – from QSPR and QSAR to direct relations with the properties of molecules and polymers via appropriate dynamic models. A special attention here should also be paid to the Kirchhoff number, which is the sum of *resistance distances* [149], calculated for any pair of vertices in the graph proceeding from the electric resistance rules of Kirchhoff. For acyclic graphs, resistance distances coincide with the graph distances and, therefore, the Wiener number may be regarded as a specific case of the more general Kirchhoff number.

This raises important questions: “Why are the variable and mostly noninteger resistance distances more general than the constant integer graph distances? What have the electric networks rules to do with graphs and molecules?” Some light on the first question may be shed proceeding from the nonreductionist complexity theory, which demands description of the systems in their entirety, not just as by summing contributions of system components. The Kirchhoff number, whose value for the distance between two vertices depends on the connectivity of the entire structure, seems to be the only topological index completely satisfying this requirement. The second question should perhaps be related to the more fundamental question for topology of space–time, as a prerequisite of any physical explanation of phenomena, including quantum mechanical one. The recent topological theory of discrete space–time [150] might mark the beginning of a development toward this goal.

4.7 Other Molecular Topology Studies

4.7.1 Theorems for the Wiener number semi–additivity [151]

In 1986, an idea of Oskar Polansky turned into a joint publication [151] on calculating the Wiener number of any graph G, $W(G)$, from the Wiener numbers $W(G_i)$ of its subgraphs $G_i \in G$. The formulae derived find application for fast calculations with combinatorial libraries [152]. The approach is based on "covering" graphs upon a common subgraph, which can be a vertex, an edge, a path or two disjoined edges. An illustration is given in Figure 11 with the simplest case of covering upon a vertex; with $d\,(u \in G)$ and $d\,(u_i \in G_i)$ being the distance numbers of the common vertex u in graph G and in its i^{th} component G_i.

$$W(G) = \sum_i W(G_i) + nd(u \in G) - \sum_i n_i d(u_i \in G_i) \qquad (4.63)$$

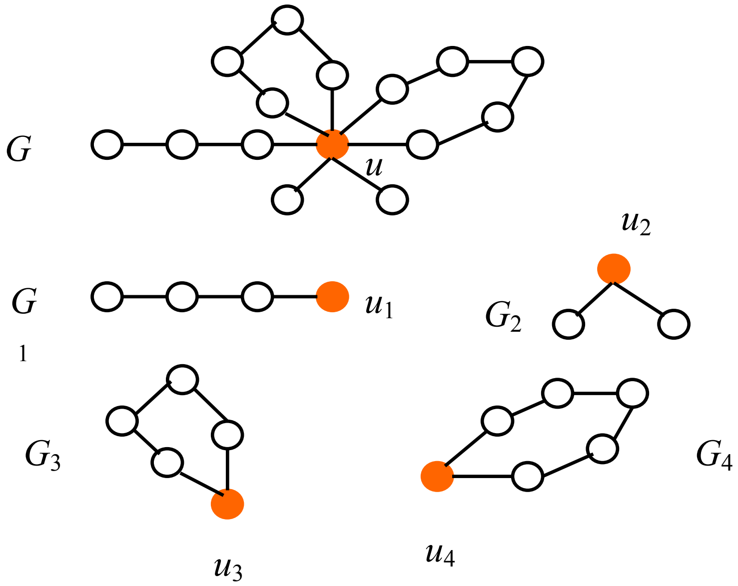


Figure 11. Illustration of the covering of subgraphs in a vertex, a procedure used to prove the theorem for calculating the Wiener number of a graph from the respective Wiener numbers of its subgraphs.

Other formulae showing how to calculate the Wiener number of a graph from covering upon an edge, a path or two disjoined edges can be found in [151].

4.7.2 Topological atomic charges and bond orders

An intriguing result was obtained (although published with some delay [153]) during my four–month visit to Monty Kier in the year 1990. By counting the self–returning walks SRW_i^k of order k that start and end in atom i, we found for a large series of compounds that the relative atomic moments of energy, f_i^k, have a limit f_i equal to the respective squared atomic coefficient in the principal eigenvector c_{1i}^2:

$$f_i = \lim_{k>>1} f_i^k = \lim_{k>>1} \left(\frac{SRW_i^k}{SRW^k} \right) = c_{1i}^2 \tag{4.64}$$

Later, this result was rigorously proved [154,155]. Thus, the relative atomic moments appear as partial atomic charges; they express the one–electron charge distribution on atoms produced by the principal eigenvector. This result was interpreted as showing that each atomic self–returning walk is associated with electron motion near the nucleus of the atom. It was also considered reminiscent of the Feynman path integral method, which accounts for all electron paths [156]. This conclusion was confirmed by the work of Nagao, Nishikawa and Aono [157], who applied the path integral method to the Hückel model and reproduced known results for the number of self–returning walks in linear and monocyclic graphs. On this basis, we concluded that each self–returning walk in molecular graph corresponds to an individual Feynman's electron trajectory [158]. Similar interpretation was made for the open–end walks, which also produce path integral contributions. We thus generalized that *molecular graphs encode the averaged information on all electron Feynman's trajectories in the molecule*. This might be a possible explanation of why molecular graphs describe so well molecular properties. The study [158], performed in collaboration with Katya Gordeeva, contained also theorems, corollaries, and relationships for the atomic and bond moments of energy in molecules and solids. Topological atomic charges, valencies, and bond orders were introduced from the limit distributions of the self–returning walks. Topological atomic and bond reactivity rules were formulated, which account for the alternating positive and negative contributions to the atomic and bond moments, shedding thus some light on the topological basis of chemical reactivity. The topological atomic charges introduced were found to correlate well with the CNDO/2 charges on carbon atoms in alkanes [159]. These charges were also shown to mirror closely the patterns of the topological charge stabilization principle of Gimarc [161] in heterocyclic conjugated compounds. Topological atomic charges were recently modified and used along with other modified topological indices (molecular connectivity, extended connectivity, overall connectivity, and the Zagreb indices) as a basis for generating 2D–chirality descriptors (joint study with Alexander Tropsha and Alexander Golbraikh [162, 163]). The modified indices included a chirality correction, which was added to the vertex degrees of asymmetric atoms in molecular graphs. The modification enabled circumventing the inability of conventional topological descriptors to distinguish chiral or enantiomeric isomers, which has been a major drawback as compared to 3D–descriptors of chemical structure.

4.7.3 The hierarchically ordered extended connectivity (HOC) concept [164–170]

This series of studies, performed in the mid 1970s with Sandy Balaban and Ovanes Mekenyan, was initially developed for the purpose of unique coding of molecular graphs and unique numbering of atoms (vertices) in them. The new concept improved the popular Morgan extended connectivity algorithm [171] by adding supplementary discriminating factors in a hierarchical iterative manner. Four versions of the HOC–procedure were proposed for application to molecules of increasing level of complexity [164–166]. The HOC–1 procedure reassigned the rank of each vertex at the k^{th} iteration, according to the sum of the $(k–1)^{\text{th}}$ ranks of the respective adjacent vertices. Several additional criteria, such as the number of cycles to which the vertex belongs, and the complete vertex degree distribution for each graph cycle, are used for more complex structures. Mathematical proof was presented for sufficiency of the various HOC procedures for vertex canonical numbering and graph orbit finding. An illustration of the HOC–1 algorithm is shown in Figure 12.

Besides for the purposes of atoms canonical numbering and molecules coding, the hierarchical extended connectivity (HOC) concept found also application for recognition of graph isomorphism and graph symmetries [167], as well as for assessing graph similarity, ordering of graphs, and defining of new topological indices [168]. The HOC–ordering of atoms was found in excellent agreement with that of the experimentally determined ^{1}H NMR chemical shifts of polycyclic benzenoid hydrocarbons having two to five fused rings [169]. Predictions were made on this basis for more than 200 proton chemical–shift orderings in 21 hexacyclic benzenoid hydrocarbons [170].

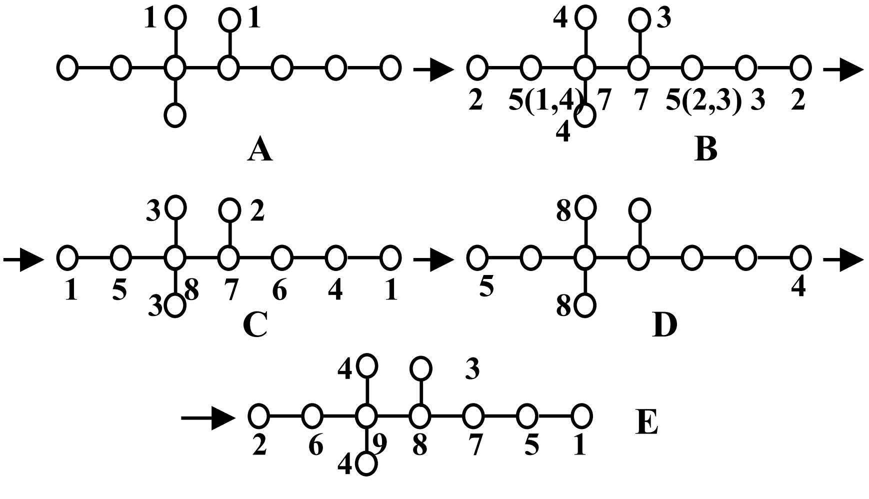


Figure 12. The HOC canonical vertex numbering: **A**) Vertices are ranked first by their degrees; **B**) Calculated extended connectivities; the two values *EC* = 5 differ by their summands; **C**) Reranking with additional differentiation within the classes of the first ranks; **D**) Recalculating the *EC*s of vertices with equal second ranks; **E**) Final ranks.

4.8 Applications of Molecular Topology to QSPR and QSAR

4.8.1 Quantitative structure–property relationships (QSPR)

The studies on the application of molecular topology to modeling properties of chemical compounds focused mainly on going beyond structure–property correlations by identifying and characterizing topological patterns in these properties. This line produced fruitful results from the very beginning, when a large number of physicochemical properties of alkanes have been shown to follow our rules of molecular branching [6]. The studies on characteristic patterns in cyclic compounds continued along the same line of trends in physicochemical properties, bringing also useful relationships with the electronic characteristics of conjugated benzenoids [44,71–76,172]. A possible generalization of the Wiener–number–based branching and cyclicity patterns was sought in the more stringent criterion requiring several graph invariants to vary in a similar manner. The concept for *comparability graph CG* was thus devised and tested with a larger number of properties [173–176]. The *CG* is a graph the vertices of which represent molecules of isomeric compounds, whereas a directed edge between two vertices stands for an isomerization transformation, associated with the same type of change (increase or decrease) for all topological indices used in the graph construction. We selected four representative topological distances: the total graph distance (the Wiener number), Randić's molecular connectivity index [43], Hosoya's non–adjacency number [3], and the number of self–returning walks [153]. The isomers located on the same path in the comparability graph are regarded comparable, because they show a stable pattern of change of these four basic graph–invariants (Figure 13).

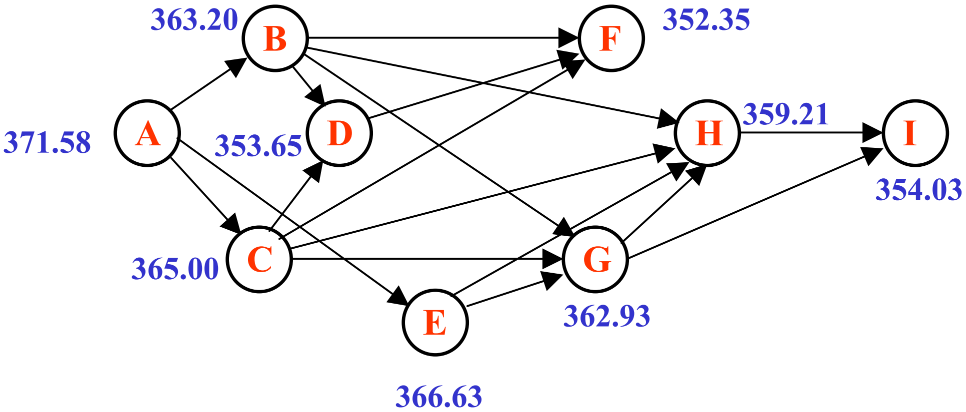


Figure 13. The boiling point subgraph of the comparability graph of C7 alkane isomers. The compounds located on the same path are regarded as an optimized correlation sample, because they obey the same topological patterns: a decrease in the Wiener number and an increase in the Randić molecular connectivity index, the Hosoya index, and the number of fourth–order self–returning walks.

When a correlation sample is composed only of such comparable isomers, it is predicted to show a high correlation with a variety of molecular properties. That was demonstrated by modeling seven thermodynamic properties of C9 and C10 alkanes [175] and spectral properties of benzenoid hydrocarbons [173,176]. The limitation to deal only with isomeric compounds was removed in the last paper of the series [176] by constructing a supercomparability graph, the number of vertices in which may vary in a broad range. Thus, our dynamic comparability approach proved to be a promising method for constructing optimized compound samples for quantitative structure–property or structure–activity models.

Another methodological question related to QSPR/QSAR was related to the manner in which a new topological index has to be tested before being used for property/activity correlations. Usually, testing against a series of isomers is preferred in order to eliminate the size effect. However, while this is a good test for *sensitivity* of the examined new descriptor toward subtle topological features of molecules, it has never been proved that topological indices that are good for isomers will have a good *predictive* power for *nonisomeric* compounds. Moreover, there were indications of the opposite [177]. The conference organized in occasion of the 25th anniversary of molecular connectivity concept of Kier and Hall [115,116] was a good opportunity to investigate this problem [114]. The best five–parameter models derived in parallel for boiling points and molar volumes of C8 alkanes and for C3–C8 ones were extrapolated to the 35 nonanes. The standard deviation of the best C8 model is only 1.14ºC vs. 1.74ºC for the best C3–C8 one. However, the extrapolation from the C8 model to nonanes resulted in a disastrous standard deviation of 6.00ºC, whereas that of the C3–C8 model, although increased to 2.08ºC, was very close to the 2.03, the value obtained for the best C3–C9 model. Even worse was the result for molar volumes, for which the C8 → C9 prediction had a standard deviation of 5.70 l/mol versus 0.62 l/mol for the C3–C8 → C9 prediction, which again is close to the best C3–C9 model value of 0.52 l/mol.

Another methodological problem, related to the same conference, questioned a key element of molecular connectivity paradigm: the lack of any prove that the mathematical function, introduced by Randić [43], $\sum_{all\ edges} 1/\sqrt{a_i a_j}$, is the best function for the molecular connectivity series of topological descriptors of Kier and Hall. The idea to use a product of the vertex degrees for each atomic bond is very logical; the question was only whether the reciprocal square root is the function that provides structure–property models with the least standard deviation. To verify that, a comparative study was performed between the Randić function, its predecessor, the second Zagreb index, $M_2 = \sum_{all\ edges} a_i a_j$, the intermediate function $ON = \sum_{all\ edges} 1/a_i a_j$ (later renamed as "modified Zagreb index" [178,179]), and the simple summation of vertex degrees, $\sum_i a_i$, which was the basis of our overall connectivity series of indices [108–111]. The comparison comprised the best linear 1– to 5– parameter models of ten physicochemical properties of C3–C8 alkane compounds. It was shown [114] that the performance of the inverse–square–root connectivity function is about the same as that of the simple inverse function, and slightly worse than that of the direct proportionality to the $(a_i a_j)$ terms. The best performance was shown by the overall connectivity models, which are based on sums of vertex degrees. These models produced the smallest standard deviations for nine of the ten examined properties. This result demonstrated the potential of the overall connectivity indices as molecular topology descriptors.

4.8.2 QSAR (Quantitative Structure–Activity Relationships) studies [46–51,54,180–187]

An idea of Ovanes Mekenyan has led to a study on the possible types of electronic drug–receptor interactions [180]. The latter were classified according to the number of reaction centers, the presence and number of parallel and consecutive steps, rate ratios of the target and side reactions, etc. It was shown that the increase in interaction complexity, due to the larger number of competing side reactions or to the larger number of receptor sites involved, increases the probability of finding correlations with extremal values of biological activity.

The major effort undertaken in the Bourgas Laboratory of Mathematical Chemistry and Chemical Informatics in the area of QSAR studies was the OASIS concept and software for multivariate regression analysis, initiated and developed with great energy by Ovanes Mekenyan [54,47,51,181]. The approach was applied in a series of joint studies with Nenad Trinajstić [182–185], Christiane Mercier and Jacque–Emil Dubois [48,186], as well as in the studies performed during my two years work at the M. D. Anderson Cancer Center in Houston [49,187].

OASIS is an abbreviation for Optimal Approach based on Structural Indices Set. It might be regarded as a second generation QSAR methodology, which goes beyond the classical Hansch approach [188] in dealing with the molecule as a whole, but not with selected atomic groups within it. This is essentially nonreductionist approach, which characterizes quantitatively molecules in three complementary levels: (*i*) molecular topology, described by graph theoretical descriptors, (*ii*) molecular 3D–geometry, and (*iii*) quantum chemistry. The arsenal of topological indices used included the Wiener index [3,189], the Randić connectivity index [43] and its generalization in the Kier and Hall series of molecular connectivity and valence connectivity indices [115,116], the Hosoya nonadjacency number [4], the Zagreb indices [65], the centric indices [81,190], Balaban's distance connectivity index [191], electropy [192], as well as several of the information theoretic indices we developed [6,37,42]. The geometric 3D–descriptors were the Wiener number 3D–analog and its information theoretic version [54], the largest interatomic distance, and some characteristic interatomic Euclidean distances. Several physicochemical properties of molecules, such as molecular volume, molecular refraction, and the *n*–octanol/water partition coefficient, were borrowed from the Hansch and Leo method [188]. The quantum mechanical SCF electronic structure in the molecules ground state was characterized using semiempirical methods like PM3, AM1, and others, proceeding from optimized molecular geometry. Later important improvement of the approach, developed by Ovanes Mekenyan's Laboratory in Bourgas included accounting for up to several hundred molecular conformations in building the QSAR models [193]. The software developed in this laboratory has found a broad application for assessing environmental toxicities [194].

5 TOPOLOGY OF COMPLEX CHEMICAL REACTIONS

5.1 Kinetic Graphs

Complex chemical reactions are described as a set of elementary steps, which form reaction routes, and are related by common substrates, products, and intermediates. The manner in which these reaction species are interrelated determines the topology of chemical reaction. There are numerous approaches to apply graph theory to the mechanistic description of chemical reactions [195–197]. My 15 years collaborative work with Oleg Temkin and his coworkers from the Moscow Institute for Fine Chemical Technology concentrated on those chemical reactions the steady–state rate laws of which can be directly derived from some invariants of specifically constructed graphs [198–206]. This class of catalytic and enzymatic reactions are characterized by a *linear reaction mechanisms*, the elementary steps of which incorporate one intermediate on both the left– and right–hand sides. We called the graphs depicting such a linear mechanism *kinetic graphs* (*KG*s). The vertices in these cyclic graphs represent only the reaction intermediates, while edges stand for their interconversions [207]. The specific version of the method, developed by M. I. Temkin [208], allows incorporating steps that do not contain intermediates by introducing “zero–intermediates” as graph vertices. An essential part of the Horiuti–Temkin [208,209] method is the concept of *reaction route*, a subset of steps whose summing cancels all intermediates. In kinetic graphs, each route is presented by a cycle. The method is illustrated below by the catalytic reaction of methanol synthesis, which is regarded as a two–route reaction with five reaction steps and four intermediates. The catalyst is denoted by *Z*, whereas the intermediates it forms include dots to symbolize the formation of complexes with the reaction species.

$$\mathrm{Z{\cdot}H_2O + CO_2} \overset{1}{\Leftrightarrow} \mathrm{Z{\cdot}H_2O{\cdot}CO_2}$$

$$\mathrm{Z{\cdot}H_2O{\cdot}CO_2} \overset{2}{\Leftrightarrow} \mathrm{Z{\cdot}CO_2 + H_2O}$$

$$\mathrm{Z{\cdot}CO_2 + H_2} \overset{3}{\Leftrightarrow} \mathrm{Z{\cdot}CO_2{\cdot}H_2}$$

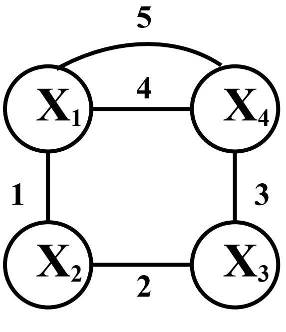
$$\mathrm{Z{\cdot}H_2O{\cdot}H_2 + 2H_2} \overset{4}{\Leftrightarrow} \mathrm{Z{\cdot}H_2O + CH_3OH}$$

$$\mathrm{Z{\cdot}\ CO_2{\cdot}H_2} \overset{5}{\Leftrightarrow} \mathrm{Z{\cdot}H_2O + CO}$$

$$\mathrm{CO_2 + 3H_2} \Leftrightarrow \mathrm{CH_3OH + H_2O}$$

$$\mathrm{CO + 2H_2} \Leftrightarrow \mathrm{CH_3OH}$$

The reaction is thus represented by a kinetic graph having four vertices, five edges, and two cycles:



Our studies [95,198–206] have shown that graph theory provides the necessary tools for classifying, coding and enumerating the complex chemical reactions having a linear mechanism, proceeding from the mechanism topological structure. Solving these problems was essential for the development of computer–assisted methods for the generation of mechanistic hypotheses and for their discrimination.

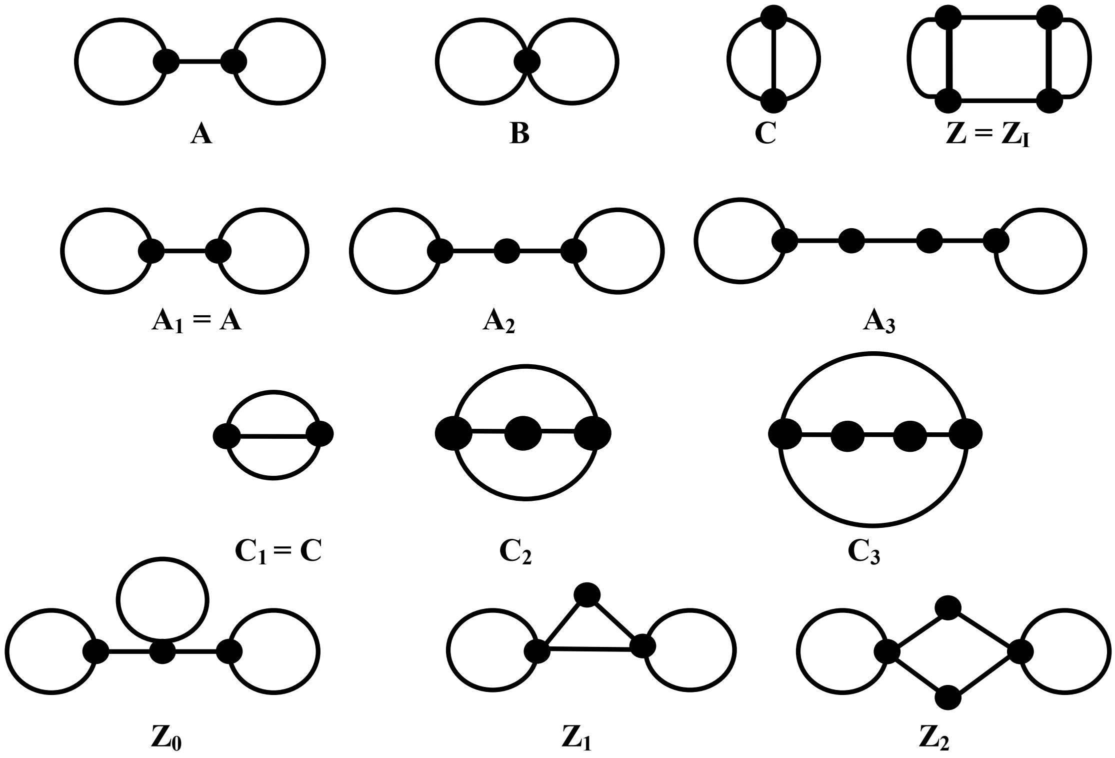


Figure 14. The four basic classes A, B, C, and D, of linear mechanisms, and their subclasses A_I, C_K, and Z_V, determined by the number of edges *I* in the bridge, the shared number of edges *K* between two cycles, and the the number of edges *V* separating two nonadjacent cycles, respectively.

5.2 Classification and Coding of Chemical Reactions with Linear Mechanisms

A hierarchical set of classification criteria was constructed:

(*i*) Number of linearly independent reaction routes (KG cycles), $M = 1,2,3,\ldots$

(*ii*) Number of intermediates (KG vertices), $N = 2,3,4,\ldots$

(*iii*) Types of interconnection of a pair of *KG* cycles (classes of two–route mechanisms; see Figure 14)

Class *A* – bridging of cycles; Class *B* – cycles sharing a common vertex

Class *C* – cycles sharing a common edge; Class *Z* – disjoint cycles (linkage via other cycles)

Prefix *n* – number of *KG* vertices with degree a ≥ 3

(*iv*) Subclasses of mechanism (number of elements connecting a pair of *KG* cycles):

Subclasses A, A_2, A_3,… (the length of the bridge, *I*)

Subclasses C, C_2, C_3,… (the number of common edges, *K*)

Subclasses Z_0, Z_1, Z_2,… (the number of edges separating a pair of cycles lacking connections of type *A*, *B* or *C*. The case $V = 0$ corresponds to a *KG* the two disjoint edges in which are actually connected by a bridge, one of the vertices of which belongs to a third cycle; see Figure 14)

(*v*) The number of vertices in each cycle, N_i

The linear code based on the above classification criteria is:

$$\mathrm{M - N - n - A_I^{\,i}\, B^j\, C_K^{\,k}\, Z_V^{\,v} - N_1, N_2, \ldots, N_M}$$

The linear code given above is abbreviated; it contains *generalized classes* with superscripts showing the number of times the particular type of cycle linkage occurs. Alternatively, one can use *specific classes*, which list all pairwise cycle linkages, e.g., ACZ^2CZ, B^2ZC^3, A^2BZAZ, etc. The mechanisms containing irreversible steps are presented by directed kinetic graphs. In such cases, the code is supplemented by the list of all edges E_i given with their type: *i*, $\bar{\imath}$ and *e*, for forward direction, reverse direction, and both directions, respectively. When the *KG* incorporates terminal (pendant) vertices, the code also includes the total number of these vertices, N_p, and the locations n_i of the vertices to which the pendant vertices *P* are attached. Thus, the code of a linear mechanism containing irreversible steps and pendant vertices becomes

$$\mathrm{M - N - n - A_I^{\,i}\, B^j\, C_K^{\,k}\, Z_V^{\,v} - N_1, N_2, \ldots, N_M - E_1, E_2, \ldots, E_E - N_p\!: n_1, n_2, \ldots, n_p}$$

5.3 Enumeration of Linear Mechanisms and Their Classes

A large–scale enumeration was performed [204] of all theoretically possible distinct linear mechanisms of chemical reactions involving up to six reaction routes, and up to twelve intermediates (with the exception of cases with M = 6 and N = 11 and 12, which were demanding unreasonably high computational time). The program KING (KINetic Graphs), which generates exhaustively all nonredundant KGs for a given number of cycles and vertices, was written by Ekaterina Gordeeva, a former coworker of the Laboratory of Mathematical Chemistry of Acad. Zefirov from the Moscow State University. The results obtained are summarized in Tables 3 and 4.

Table 3. Total number of kinetic graphs with M = 2–6 reaction routes and N = 2–12 intermediates

M \ N	2	3	4	5	6	7	8	9	10	11	12
2	1	2	4	7	10	14	19	24	30	37	44
3	1	3	12	27	65	129	245	422	710	1113	1710
4	1	5	23	85	276	764	1935	4466	9583	19291	36859
5	1	6	43	210	924	3403	11242	33156	89789	224621	526346
6	1	8	72	469	2652	12644	52727	194909	651008	CE*	CE*

*Combinatorial explosion

Table 4. Total number of classes of linear mechanisms having 2–6 reaction routes and 2–12 reaction intermediates

M \ N	2	3	4	5	6	7	8	9	0	11	12
2	1	1	1	0	0	0	0	0	0	0	0
3	1	2	6	3	2	1	0	0	0	0	0
4	1	4	14	24	33	19	11	4	1	0	0
5	1	5	30	85	192	249	250	153	77	26	7
6	1	7	55	239	798	1746	2800	3082	2576	CE	CE

It was found that, at a constant number of reaction routes and an increasing number intermediates, the number of classes passes through a maximum and behaves close to the normal distribution. The very high number of classes and mechanisms found far exceeded some previously published estimates. This indicates that enumerations based on mechanistic chemical information only, do not take into account the very high degree of variety of mechanisms topological structure. On the other hand, it also signals that some topologically allowed mechanisms could not be chemically possible. More studies are needed for the ultimate elucidation of the problem. Other unanswered problems remain with the inclusion of irreversible and equilibrium steps, which in graph theoretical terms is equivalent to a transition from undirected graphs to directed graphs and such with pendant vertices.

5.4 Complexity of Linear Reaction Mechanisms

Early in the development of our graph theoretical approach to chemical reactions networks, we felt the necessity to introduce a quantitative estimate of mechanistic complexity [95,202,203,207]. Such a measure would be beneficial for the process of (*i*) hypothesis formulation, assuming some upper limit above which the hypotheses would become unreasonably complex; (*ii*) selecting the simplest mechanism in situations when several mechanisms fit equally well the experimental data; (*iii*) constructive enumeration of kinetic graphs. We constructed a complexity measure that mirrors both levels of mechanistic complexity – the complex topology of the kinetic graph, and the complexity of the derived kinetic model. The complexity index K developed for mechanisms containing only reversible steps has the following form:

$$K = MN(N-1)T_i + 2N\sum_{p=1}^{M}\sum_{k=0}^{k(\max)} D_{pk} \tag{5.1}$$

Here, T_i is the number of spanning trees having a root at vertex *i*. D_{pk} is termed algebraic complement of cycle *p*. It is the base determinant of the subgraph resulting from contracting the *pk* cycle to a single vertex. The double sum counts the number of spanning trees of the KG subgraphs obtained after each such *pk* cycle contracting for each of the *p* cycles. Since the calculation of *K* depends heavily on the number of spanning trees, we derived general formulas for T_i of linear networks having one, two, three, and four reaction routes. This helped in calculating the complexity level of all topologically distinct mechanisms having two to six reaction routes and 2 to 12 intermediates. Eq. (5.1) also helped in identifying the trends that increase mechanistic complexity. They are listed below in a hierarchical manner:

i) The strongest complexity factor is the number of reaction routes *M*.

ii) At a constant number of routes *M*, the number of intermediates *N* is a stronger complexity factor than the route connectedness and the intermediate connectedness for networks having 2 to 4 intermediates. However, for *N* = 5, the two factors are comparable, whereas for $N \geq 6$ the enhanced connectedness prevails.

iii) At a constant number of routes and intermediates, the enhancing of the route interconnectedness increases network complexity when the additional cycle connection is of class C (common reaction step), but keep constant for classes A and B connectedness.

iv) At a constant number of routes and intermediates, and a constant number of route interconnections, complexity index increases with a rate following the relationship:

A < B < C.

v) At a constant number of routes and intermediates, and a constant type and class, the complexity index increases depending on the number of reaction steps connecting the two routes, according to the inequalities: $\mathbf{A_3 < A_2 < A}$, and $\mathbf{C < C_2 < C_3}$.

The complexity analysis also revealed classes of transformations, which do not change the KG complexity. All isocomplexity relationships found for a given number of routes were summarized in flowcharts (see Figure 15). The classes with the same complexity are located in the flowcharts along vertical lines, whereas horizontal or diagonal lines show the relations of increasing complexity.

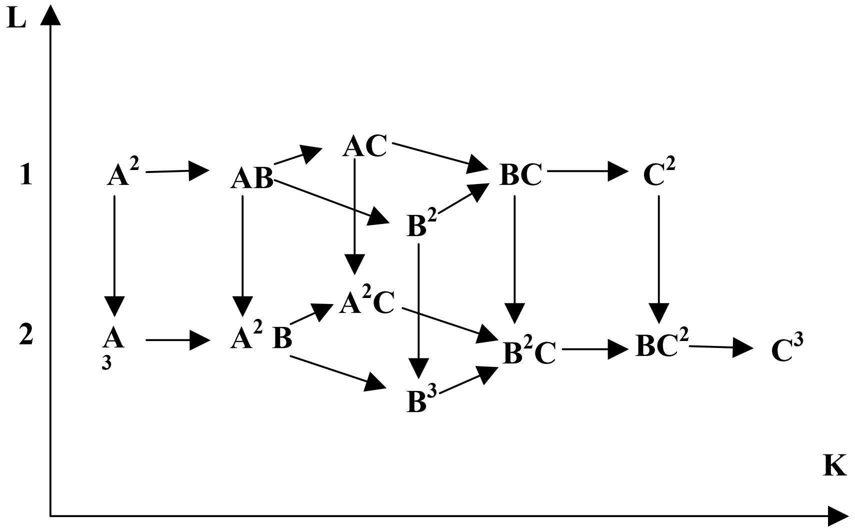


Figure 15. Complexity flowchart for types *L* and classes of three–route reaction networks with linear mechanism. Complexity increases along horizontal and diagonal lines and keeps constant along vertical lines. The type index *L* shows the number of interconnections between the three routes.

The possibilities for extending the topological approach from reaction networks with linear mechanisms to those with nonlinear mechanisms were discussed in our monograph [197]. Other general problems of chemical reaction networks were discussed in that book, such as dealing with chemical reaction as a combinatorial object, classification of reaction mechanisms based on bipartite graphs, and the relation between the topological structure of a mechanism and the structure of its kinetic model.

6 SOME TOUGHTS ON THE FUTURE DEVELOPMENTS IN MATHEMATICAL CHEMISTRY

The variety of areas and problems in discrete applied mathematics (chemical graph theory and chemical information theory) discussed in this article illustrates many of the main avenues of this relatively young branch of theoretical chemistry. Focused on characterizing chemical structure and the relation between structure and properties, mathematical chemistry obtained important results and developed powerful tools, which helped in the design of new chemical products and drugs. Topological indices are invaluable for fast data mining of combinatorial libraries of chemical compounds. After looking back one has to turn to the future and ask: “What lies ahead: variations on the same theme or radically new developments?” The recent explosive development of mathematical biology and bioinformatics offers a golden opportunity for potential applications of the accumulated knowledge in mathematical chemistry. Biological and ecological networks manifest topological patterns similar to those of atomic and molecular networks. The gigantic task of explaining the specific biological functions proceeding from the structure of cellular protein–protein interaction networks, metabolic, and gene regulatory networks, might be regarded as a natural extension of what we have learned about molecular topology and the topology of chemical reaction networks. Chemical graph theory could only be enriched from the interaction with this new area of research, which attracts many mathematicians, physicists and chemists. However, there is a major difference between the way graph theory is applied to molecules and to a cellular network. The latter is a dynamic, evolving network, whereas molecular structure is regarded as static. The lack of *dynamic* molecular graph theory limits the possibilities for deriving molecular properties from molecular topology. Statistical regression models result but the next step – from correlation to functional topology–property relationships, based on a certain mechanistic dynamic model – is still to be done. My recent work [143] on deriving relationships between the Wiener and Kirchhoff topological indices and the radius of gyration and viscosity of polymers within the dynamic Rouse – Zimm model might be regarded as a step in this direction.

7 REFERENCES

[1] L. Brillouin, *Science and Information Theory*, Academic Press, New York, 1956.
[2] C. E. Shannon and W. Weaver, *Mathematical Theory of Communications*. University of Illinois Press, Urbana IL, 1949.
[3] H. Wiener, Structural Determination of Paraffin Boiling Points, *J. Am. Chem. Soc.* **1947**, *69*, 17–20; Relation of the Physical Properties of the Isomeric Alkanes to Molecular Structure, *J. Phys. Chem.* **1948**, *52*, 1082–1089.
[4] H. Hosoya, A Newly Proposed Quantity Characterizing Topological Nature of Structural Isomers of Saturated Hydrocarbons, *Bull. Chem. Soc. Jpn.* **1971**, *44*, 2332–2339.
[5] I. Gutman and N. Trinajstić, Graph Theory and Molecular Orbitals, *Topics Curr. Chem.* **1973**, *42*, 49–93.
[6] D. Bonchev and N. Trinajstić, Information Theory, Distance Matrix, and Molecular Branching, *J. Chem. Phys.* **1977**, *67*, 4517–4533.
[7] D. Bonchev, On the Dependence of the Fine Structure Constant on Time, *God. Vissh Khim.–Technol Inst. Bourgas*, **1966**, *3*, 89–98.
[8] P. A. M. Dirac, The Cosmical Constants, *Nature* 139, 323, 1937.

[9] E. R. Harrison, *Nature* 197, 1257, 1963.
[10] D. Bonchev and O. Mekenyan, An Estimate of the Upper Limit of the Periodic Table Based on the Eddington Number, *Rev. Roum. Chim.* **1981**, *26*, 349–357.
[11] D. Bonchev, On the Nature and Motions of Electron. **1997**, unpublished.
[12] A. Mowshowitz, Entropy and the Complexity of Graphs. I. An Index of the Relative Complexity of a Graph, *Bull. Math. Biophys.* **1968**, *30*, 175–224.
[13] D. Bonchev and T. Peev, Information Theoretic Study of Chemical Elements. Mean Information Content of a Chemical Element, *God. Vissh Khim.–Technol Inst. Bourgas* **1973**, *10*, 561–574.
[14] D. Bonchev, C. Tashkova, and R. Ljuzkanova, On the Correlation Between Enthalpy of Formation, Atomic Number, and Information Content of Alkali Halides, *Dokl. BAN* **1975**, *28*, 225–228.
[15] D. Bonchev and V. Kamenska, D. Kamenski, Informationsgehalt Chemischer Elemente, *Monatsh. Chem.* **1977**, *108*, 477–487.
[16] D. Bonchev and V. Kamenska, Informationscharacteristiken der Perioden und Unterperioden im Periodensystem, *Monatsh. Chem.* **1978**, *109*, 551–556.
[17] D. Bonchev and V. Kamenska, Information Theory in Describing the Electronic Structure of Atoms, *Croat. Chem. Acta* **1978**, *51*, 19–27.
[18] D. Bonchev, Information Indices for Atoms and Molecules, *Commun. Math. Comput. Chem.* (*MATCH*) **1979**, *7*, 65–113.
[19] D. Bonchev, V. Kamenska, and C. Tashkova, Equations for the Elements in the Periodic Table Based on Information Theory, *Commun. Math. Comput. Chem.* (*MATCH*) **1976**, *2*, 117–122.
[20] D. Dimov and D. Bonchev, Spin–Information Equations of the Groups and Periods in the Periodic Table of Chemical Elements, *Commun. Math. Comput. Chem.* (*MATCH*) **1976**, *2*, 111–115.
[21] D. Bonchev, Information Theory Interpretation of the Pauli Principle and Hund Rule, *Intern. J. Quantum Chem.* **1981**, *19*, 673–679.
[22] D. Bonchev and V. Kamenska, Electronenübergänge und die Veränderung des Informations–Gehaltes Chemicher Elemente, *Monatsh. Chem.* **1979**, *110*, 607–612.
[23] D. Bonchev and V. Kamenska, Correlations Between Information Indices and Properties of Chemical Elements, *Commun. Math. Comput. Chem.* (*MATCH*) **1979**, *7*, 113–133.
[24] The Second Hary Wiener International Memorial Conference "The Periodic Table: Into the 21st Century", July 14–20, 2003, Kananaskis, Banff, Alberta, Canada.
[25] D. Bonchev, Periodicity of Chemical Elements and Nuclides. Information Theoretic Analysis; in: *Mathematics of the Periodic Table*, Eds. D. H. Rouvray and R. B. King, Nova Science, New York, in press.
[26] D. Bonchev and V. Kamenska, Predicting the Properties of the 113–120 Transactinide Elements, *J. Phys. Chem.* **1981**, *85*, 1177–1186.
[27] T. Peev, D. Bonchev, B. Rousseva, and A. Dimitrov, Information Studies of 2ß–Stable Atomic Nuclei, *God. Vissh. Techn. High. Sch. Phys.* (Sofia) **1974**, *9*, 73–86.
[28] T. Peev, D. Bonchev, A. Dimitrov, and B. Rousseva, Differential Information Characteristics of 2ß–Stable Isotopes, *God. Vissh. Techn. High. Sch. Phys.* (Sofia) **1977**, *11*, 87–98.
[29] D. Bonchev, T. Peev, and B. Rousseva, Information Study of Atomic Nuclei. Information on Proton–Neutron Composition, *Commun. Math. Comput. Chem.* (*MATCH*) **1976**, *2*, 123–137.
[30] B. Rousseva and D. Bonchev, A Theoretic–Information Variant of Nuclide Systematics, *Commun. Math. Comput. Chem.* (*MATCH*) **1978**, *4*, 173–192.
[31] B. Rousseva and D. Bonchev, Empirical Correlations Between the Nuclear Binding Energy and Electron Energy of Isodifferent Nuclides, *Radiochem. Radioanal. Lett.* **1979**, *40*, 41–50.
[32] B. Rousseva and D. Bonchev, Prediction of the Nuclear Binding Energies of the Nuclides of Period VII, *Radiochem. Radioanal. Lett.* **1980**, *45*, 341–346.
[33] S. M. Dancoff and H. Quastler, The Information Content and Error Rate of Living Things; in: *Essays on the Use of Information Theory in Biology*, Ed. H. Quastler, University of Illinois, Urbana, 1953.
[34] H. Branson; in: *Essays on the Use of Information Theory in Biology*, Ed. H. Quastler, University of Illinois, Urbana, 1953.
[35] N. Rashewsky, Life, Information Theory, and Topology, *Bull. Math. Biophys.* **1955**, *17*, 229–235.
[36] E. Trucco, A Note on the Information Content of Graphs, *Bull. Math. Biophys.* **1956**, *18*, 129–135; E. Trucco, On the Information Content of Graphs: Compound Symbols; Different States for Each Point, *ibid.* **1956**, *18*, 237–253.
[37] D. Bonchev, O. Mekenyan, and N. Trinajstić, Isomer Discrimination by Topological Information Approach, *J. Comput. Chem.* **1981**, *2*, 127–148.
[38] A. Mowshowitz, Entropy and the Complexity of Graphs: IV. Entropy Measures and Graphical Structure, *Bull. Math. Biophys.* **1968**, *30*, 533–546.
[39] Bonchev, D. Kamenski, and V. Kamenska, Symmetry and Information Content of Chemical Structures, *Bull. Math. Biol.* **1976**, *38*,119–133.

[40] D. Bonchev, A. T. Balaban and O. Mekenyan, Generalization of the Graph Center Concept, and Derived Topological Indexes, *J. Chem. Inf. Comput. Sci.* **1980**, *20*,106–113.
[41] S. H. Bertz, The First General Index of Molecular Complexity, *J. Am. Chem. Soc.* **1981**, *103*, 3599–3601.
[42] D. Bonchev, *Information–Theoretic Indices for Characterization of Chemical Structures*. Research Studies Press, Chichester, U. K., 1983.
[43] M. Randić, On Characterization of Molecular Branching, *J. Am. Chem. Soc.***1975**, *97*, 6609–6615.
[44] O. Mekenyan, D. Bonchev, and N. Trinajstić, Chemical Graph Theory Modeling the Thermodynamic Properties of Molecules, *Intern. J. Quantum Chem. Symp.* **1980**, *18*, 369–380.
[45] D. Papazova, N. Dimov, and D. Bonchev, Calculation of Gas Chromatographic Retention Indices of Isoalkanes Based on a Topological Approach, *J. Chromatogr.* **1980**, *188*, 297–303.
[46] O. Mekenyan, D. Peitchev, D. Bonchev, N. Trinajstic, and I. Bangov, Modeling the Interaction of Small Organic Molecules with Biomacromolecules. I. Interaction of Substituted Pyridines with anti–3–Azopyridine Antibody, *Arzneim. Forsch.* **1986**, *36*, 176–183.
[47] O. Mekenyan and D. Bonchev, OASIS Method for Predicting Biological Activity of Chemical Compounds, *Acta Pharm. Jugosl.* **1986**, *36*, 225–237.
[48] C. Mercier, O. Mekenyan, J.–E. Dubois, and D. Bonchev, DARC/PELCO and OASIS Methods. I. Metodological Comparison. Modeling Purines pKa and Antitumor Activity, *Eur. J. Med. Chem.* **1991**, *26*, 575–592.
[49] D. Bonchev, C. F. Mountain, W.A. Seitz, A.T. Balaban, Modeling the Anticancer Action of Some Retinoid Compounds by Making Use of the OASIS Method, *J. Med. Chem.* **1993**, *36*, 1562–1569.
[50] D. Bonchev, C. F. Mountain, and W. A. Seitz, On the Conformational Flexibility of Aromatic Retinoids, *Biophys. Chem.* **1994**, *51*, 59–69.
[51] O. Mekenyan, S. Karabunarliev, and D. Bonchev, The Microcomputer OASIS System for Predicting the Biological Activity of Chemical Compounds, *Comput. Chem.* **1990**, *14*, 193–200.
[52] CODESSA 2.13, Semichem, 7204 Mullen, Shawnee, KS 66216, U.S.A., www: http://www.semichem.com
[53] DRAGON PROFESSIONAL and DRAGON PLUS version 5, Piazza Repubblica, 32 – 20124 Milano, Italy, Web Site: http://www.talete.mi.it.
[54] O. Mekenyan and D. Bonchev, Molecular Structure Descriptors and Biological Activity; in: FECS Third International Conference on Chemistry and Biotechnology of Biologically Active Natural Products, Sept. 16–21, 1985, Sofia, Vol.3, Half–Hour Plenary Lectures, Bulg. Acad. Sci., Sofia, 1985, pp. 365–393.
[55] F. Fratev, D. Bonchev, and V. Enchev, A Theoretic–Information Approach to Ring and Total Aromaticity in Ground and Excited State, *Croat. Chem. Acta* **1980**, *53*, 545–554.
[56] O. Mekenyan, D. Bonchev, and D. N. Dimitrov, Information Theoretic Approach to the Aromaticity of Chemical Compounds Based on Superdelocalizability Indices, *Rev. Roumaine Chim.* **1991**, *36*, 363–377.
[57] R. Carbó–Dorca, D. Robert, L. Amat, X. Gironés, and E. Besalú, Molecular Quantum Similarity in QSAR and Drug Design, *Lecture Notes in Chemistry*,Vol. 73, Springer Verlag, Berlin, 2000.
[58] Y. A. Zhdanov, On the Amount of Information in the Molecules of Bioorganic Compounds, *Biofiz.* **1967**, *12*, 715–717.
[59] Y. A. Zhdanov, *Information Entropy in Organic Chemistry*, Rostov University, Rostov, Russia, 1979.
[60] D. Bonchev, Shannon's Information and Complexity: in: *Mathematical Chemistry Series*, *Volume* 7, *Complexity in Chemistry*, Eds. D. Bonchev and D. H. Rouvray, Taylor & Francis, 2003, pp. 155–187.
[61] D. Bonchev and N. Trinajstić, On Topological Characterization of Molecular Branching, *Intern. J. Quantum Chem. Symp.* **1978**, *12*, 293–303.
[62] D. Bonchev, J. V. Knop, and N. Trinajstić, Mathematical Models of Branching, *Commun. Math. Comput. Chem.* (*MATCH*) **1979**, *6*, 21–47.
[63] Bonchev, D. Topological Order in Molecules. 1. Molecular branching revisited, *J. Mol. Struct.* (*Theochem*) **1995**, *336*, 137–156.
[64] L. Lovasz and J. Pelikan, On the Eigenvalues of Trees, *Period. Math. Hung.* **1973**, *3*, 175–182.
[65] I. Gutman, B. Ruščić, N. Trinajstić, and C. F. Wilcox, Graph Theory and Molecular Orbiytals. XII. Acyclic Polyenes, *J. Chem. Phys.* **1975**, *62*, 3399–3405.
[66] S. Bertz, Branching in Graphs and Molecules, *Discr. Appl. Math.* **1988**, *19*, 65–83.
[67] M. Randić, On Molecular Branching, *Acta Chim. Slovenica* **1997**, *44*, 57–77.
[68] E. Ruch and I. Gutman, The Branching Extent of Graphs, *J. Combinatorics* **1979**, *4*, 285–295.
[69] O. E. Polansky and D. Bonchev, The Wiener number of graphs. I. General Theory and Changes Due To Graph Operations, *Commun. Math. Comput. Chem.* (*MATCH*) **1986**, *21*, 133–186
[70] O. E. Polansky and D. Bonchev, Theory of the Wiener Number of Graphs II. Transfer Graphs and Some of Their Metric Properties, *Commun. Math. Comput. Chem.* (*MATCH*) **1990**, *25*, 3–40.
[71] D. Bonchev, O. Mekenyan, J.V. Knop, and N. Trinajstić, On Characterization of Monocyclic Structures, *Croat. Chem. Acta* **1979**, *52*, 361–367.
[72] D. Bonchev, O. Mekenyan, and N.Trinajstić, Topological Characterization of Cyclic Structures, *Intern. J.*

Quantum Chem. **1980**, *17*, 845–893.

[73] O. Mekenyan, D. Bonchev, and N. Trinajstić, Topological Rules for Spirocompounds, *Commun. Math. Comput. Chem.* (*MATCH*) **1979**, *6*, 93–115.

[74] O. Mekenyan, D. Bonchev and N. Trinajstić, On Algebraic Characterization of Bridged Polycyclic Compounds, *Intern. J. Quantum Chem*. **1981**, *19*, 929–955.

[75] O. Mekenyan, D. Bonchev, and N. Trinajstić, A Topological Characterization of Cyclic Structures With Acyclic Branches, *Commun. Math. Comput. Chem.* (*MATCH*) **1981**, *11*, 145–168.

[76] O. Mekenyan, D. Bonchev, and N. Trinajstić, Structural Complexity and Molecular Properties of Cyclic Molecules With Acyclic Branches, *Croat. Chem. Acta* **1983**, *56*, 237–261.

[77] A. T. Balaban, D. Bonchev, X. Liu, and D. J. Klein, Molecular Cyclicity and Centricity of Polycyclic Graphs. I. Cyclicity Based on Resistance Distances or Reciprocal Distances, *Int. J. Quantum Chem*. **1994**, *50*, 1–20.

[78] D. J. Klein and M. Randić, Resistance Distance, *J. Math. Chem*. **1993**, *12*, 81–95.

[79] M. Randić and D. Plavšić, On the Concept of Molecular Complexity, *Croat. Chem. Acta* **2002**, *75*, 107–116.

[80] F. Harary, *Graph theory*, Addison–Wesley, Reading, MA, 1969.

[81] D. Bonchev, A. T. Balaban and O. Mekenyan, Generalization of the Graph Center Concept, and Derived Topological Indexes, *J. Chem. Inf. Comput. Sci*. **1980**, *20*, 106–113.

[82] D. Bonchev, A. T. Balaban, and M. Randić, The Graph Center Concept for Polycyclic Graphs, *Intern. J. Quantum Chem*. **1981**, *19*, 61–82.

[83] D. Bonchev, O. Mekenyan, and A. T. Balaban, An Iterative Procedure for the Generalized Graph Center in Polycyclic Graphs, *J. Chem. Inf. Comput. Sci*. **1989**, *29*, 91–97.

[84] D. Bonchev, The Concept for the Center of a Chemical Structure and Its Applications, *Theochem* **1989**, *185*, 155–168.

[85] A. T. Balaban, I. Motoc, D. Bonchev, and O. Mekenyan, Topological indices for QSAR. *Topics Curr. Chem.* **1983**, *114*, 21-56.

[86] I. Motoc, A. T. Balaban, O. Mekenyan, and D. Bonchev, Topological Indices: Inter–Relations and Composition, *Commun. Math. Comput. Chem.* (*MATCH*) **1982**, *13*, 369–404.

[87] A. Dobrinin and V. A. Skorobogatov, Metric Analysis of Graphs, *Commun. Math. Comput. Chem.* (*MATCH*) **1988**, *23*, 105–151.

[88] A. Dobrinin, Regular Graphs Having the Same Path Layer Matrix, *J. Graph Theory* **1990**, *14*, 141–148.

[89] O. Ivanciuc and D. J. Klein, Building–Block Computation of Wiener–Type Indices for the Virtual Screening of Combinatorial Libraries, *Croat. Chem. Acta* **2002**, *75*, 577–601.

[90] D. Bonchev and A. T. Balaban, Topological Centric Coding and Nomenclature of Polycyclic Compounds. I. Condensed Benzenoid Systems (Polyhexes, Fusenes), *J. Chem. Inf. Comput. Sci*. **1981**, *21*, 223–229.

[91] D. Bonchev, Principles of a Novel nomenclature of Organic Compounds, *Pure Appl. Chem*. **1983**, *55*, 221–228.

[92] A. T. Balaban, D. Bonchev, and W. A. Seitz, Topological/Chemical Distances and Graph Centers in Molecular Graphs With Multiple Bonds, *Theochem* **1992**, *280*, 253–260.

[93] D. Bonchev and A.T. Balaban, Central Vertices Versus Central Rings in Polycyclic Condensed Systems, *J. Math. Chem*. **1994**, *14*, 287–304.

[94] D. Minoli, Combinatorial Graph Complexity, *Atti. Acad. Naz. Lincei Rend*. **1976**, *59*, 651– 661.

[95] D. Bonchev, The Problems of Computing Molecular Complexity: in: *Computational Chemical Graph Theory*, Ed. D. H. Rouvray, Nova Publications, New York, 1990, pp. 34–67.

[96] D. Bonchev, O. N. Temkin, and D. Kamenski, On the Complexity of Linear Reaction Mechanisms, *React. Kinet. Catal. Lett*. **1980**, *15*, 119–124.

[97] P. E. John, R. Mallion, and I. Gutman, An Algorithm for Counting Spanning Trees in Labeled Molecular Graphs Homeomorphic to CataCondensed Systems, *J. Chem. Inf. Comput. Sci*. **1998**, *38*, 108–112.

[98] S. H. Bertz, A Mathematical Model of Molecular Complexity; in: *Chemical Applications of Topology and Graph Theory*, Ed. R. B. King, Elsevier, Amsterdam, 1983, pp. 206–221.

[99] D. Bonchev and O. E. Polansky, On the Topological Complexity of Chemical Systems; in: *Graph Theory and Topology in Chemistry*, Eds. R. B. King and D. H. Rouvray, Elsevier, Amsterdam, 1987, pp. 126–158.

[100] D. Bonchev, W. A. Seitz, The Concept of Complexity in: Chemistry, in: *Concepts in Chemistry: A Contemporary Challenge*, Ed. D. H. Rouvray, Research Studies Press, Taunton, U. K., 1996, pp. 348–376.

[101] S. H. Bertz and W. F. Wright, The Graph Theory Approach to Synthetic Analysis: Definition and Application of Molecular Complexity and Synthetic Complexity, *Graph Theory Notes New York Acad. Sci*. **1998**, *35*, 32–48.

[102] G. Rücker and C. Rücker, Walk Counts, Labyrinthicity, and Complexity of Acyclic and Cyclic Graphs and Molecules, *J. Chem. Inf. Comput. Sci*. **2000**, *40*, 99–106.

[103] M. Randić and D. Plavšić, On the Concept of Molecular Complexity, *Croat. Chem. Acta* **2002**, *75*, 107–116.

[104] S. Nikolić, N. Trinajstić, I. M. Tolić, G. Rücker, and C. Rücker, On Molecular Complexity Indices; in: *Mathematical Chemistry Series. Vol. 7*, *Complexity in Chemistry*, Eds. D. Bonchev and D. H. Rouvray, Taylor & Francis, London, 2003, pp. 29–89.

[105] D. Bonchev and D. H. Rouvray (Eds.), *Complexity in Chemistry*, *Biology*, *and Ecology*, Springer, New York, in press.
[106] M. M. Waldrop, *Complexity: The Emerging Science at the Edge of Order and Chaos*, Touchstone–Simon & Schuster, New York, 1992.
[107] G. Nikolis and I. Prigogine, *Exploring Complexity*, Freeman, New York, 1989.
[108] J. Rosen, *Symmetry Discovered: Concepts and Applications in Nature and Science*, Cambridge University Press, Cambridge, MA, 1975.
[109] D. Bonchev, Kolmogorov's Information, Shannon's entropy, and Topological Complexity of Molecules, *Bulg*. *Chem*. *Commun*. **1995**, *28*, 567–582.
[110] D. Bonchev, Novel Indices for the Topological Complexity of Molecules, *SAR QSAR Environ*. *Res*. **1997**, *7*, 23–43.
[111] D. Bonchev, Overall Connectivity and Molecular Complexity; in: *Topological Indices and Related Descriptors*, Eds. J. Devillers and A.T. Balaban, Gordon and Breach, Reading, U.K., 1999, pp. 361–401.
[112] D. Bonchev, Overall Connectivities/Topological Complexities: A New Powerful Tool for QSPR/QSAR, *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 934–941.
[113] Bertz, S. H. and Sommer, T. J. Rigorous Mathematical Approaches to Strategic Bonds and Synthetic Analysis Based on Conceptually Simple New Complexity Indices, *Chem*. *Commun*. **1997**, 2409–2410.
[114] S. H. Bertz and W. C. Herndon, The Similarity of Graphs and Molecules; in: *Artificial Intelligence Applications to Chemistry*, Eds. T. H. Pierce and B. A. Hohne, ACS, Washington, DC, 1986, pp.169–175.
[115] D. Bonchev, Overall Connectivity – A Next Generation Molecular Connectivity, *J. Mol. Graphics Model.* **2001**, *20*, 55–65.
[116] L. B. Kier and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976.
[117] L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure–Activity Analysis*, Wiley, New York, 1986.
[118] D. Bonchev, The Overall Wiener Index – A New Tool for Characterization of Molecular Topology, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 582–592.
[119] D. Bonchev and N. Trinajstić, Overall Molecular Descriptors. 3. Overall Zagreb Indices, SAR QSAR Environ. Res. **2001**, *12*, 213–235.
[120] G. Rücker and C. Rücker, Substructure, Subgraph and Walk Counts as Measures of the Complexity of Graphs and Molecules, *J. Chem. Inf. Comput. Sci.* **2001**, *41*,1457–1462.
[121] D. Bonchev, On the Complexity of Platonic Solids, *Croat*. *Chem*. *Acta* **2004**, *77*, 175–178.
[122] H. Jeong, B. Tombor, R. Albert, Z. N. Oltval, and A.–L. Barabasi, The Large–Scale Organization of Metabolic Networks, *Nature* **2000**, *407*, 651–654.
[123] D. J. Watts and S. H. Strogatz, Collective Dynamics of "Small–World" Networks, *Nature* **1998**, *393*, 440–442.
[124] D. Bonchev. Complexity of Protein–Protein Interaction Networks, Complexes and Pathways; in: *Handbook of Proteomics Methods*, Ed. M. Conn, Humana, New York, 2003, pp. 451–462.
[125] D. Bonchev, On the Complexity of Directed Biological Networks, *SAR QSAR Environ*. *Res*. **2003**, *14*, 199–214.
[126] D. Bonchev, Shannon's Information and Complexity; in: *Mathematical Chemistry Series*, *Volume* 7, *Complexity in Chemistry*, Eds. D. Bonchev and D. H. Rouvray, Taylor & Francis, 2003, pp. 155–187.
[127] D. Bonchev, Complexity Analysis of Yeast Proteome Network, *Chem*. *Biodiversity* **2004**, *1*, 312–326.
[128] D. Bonchev and G. A. Buck, Quantitative Measures of Network Complexity; in: *Complexity in Chemistry*, *Biology and Ecology*, Eds. D. Bonchev and D. H. Rouvray, Springer, New York, in press.
[129] D. Bonchev, O. Mekenyan, and H. Fritsche, An Approach to the Topological Modeling of Crystal Growth, *J*. *Cryst*. *Growth* **1980**, *49*, 90–96.
[130] D. Bonchev, O. Mekenyan, and H. Fritsche, A Topological Approach to Crystal Vacancy Studies. I. Model Crystallites with a Single Vacancy, *Phys*. *stat*. *sol*. (*a*) **1979**, *55*, 181–187.
[131] O. Mekenyan, D. Bonchev, and H. Fritsche, A Topological Approach to Crystal Vacancy Studies. II. Model Crystallites with Two and Three Vacancies, *Phys*. *Stat*. *sol*.(*a*) **1979**, *56*, 607–614.
[132] O. Mekenyan, D. Bonchev, and H. Fritsche, A Topological Approach to Crystal Defect Studies, *Z. Phys*. *Chem*. (*Leipzig*) **1984**, *265*, 959–967.
[133] H. G. Fritsche, D. Bonchev, and O. Mekenyan, Deutung der Magischen Zahlen von Argonclustern als Extremwerte Topologischer Indizes, *Z. Chem*. **1987**, *27*, 234.
[134] H. G. Fritsche, D. Bonchev, and O. Mekenyan, On the Topologies of $(M_{13})_{13}$ Superclusters of Ruthenium, Rhodium and Gold, *J. Less–Common Metals* **1988**, *141*, 137–143.
[135] H. G. Fritsche, D. Bonchev, and O. Mekenyan, Are Small Clusters of Inert–Gas Atoms Polyhedra of Minimun Surfaces? *Phys*. *Stat*. *Sol*.(*b*) **1988**, *148K*, 101–104.
[136] H. G. Fritsche, D. Bonchev, and O. Mekenyan, The Optimum Topology of Small Clusters, *Z. Phys*. *Chem*. (*Leipzig*) **1989**, *270*, 467–476.
[137] H. G. Fritsche, D. Bonchev, and O. Mekenyan, A Topological Approach to Studies of Ordered Structures of

Absorbed Gases in Host Lattices (I). The Structure of –PdD0.5, *Crystal Res. Technol.* **1983**, *18*, 1075–1081.

[138] D. Bonchev and O. Mekenyan, A Topological Approach to the Calculation of the π–Electron Energy and Energy Gap of Infinite Conjugated Polymers, *Z. Naturforsch.* **1980**, *35a*, 739–747.

[139] D. Bonchev, O. Mekenyan, and V. Kamenska, A Topological Approach to the Modeling of Polymer Properties (The TEMPO Method), *J. Math. Chem.* **1992**, *11*, 107–132.

[140] D. Bonchev, O. Mekenyan, and O. E. Polansky, A Topological Approach to the Predicting of the Electron Energy Characteristics of Conjugated Infinite Polymers. II. PPP– Calculations, *Z. Naturforsch.* **1981**, *36a*, 643– 646; III. The Influence of Some Structural Modifications of Polymers, *ibid*. 647–650.

[141] O. Mekenyan, S. Dimitrov, and D. Bonchev, Graph–Theoretical Approach to the Calculation of Physico–Chemical Properties of Polymers, *Eur. Polymer J.* **1983**, *12*, 1185–1193.

[142] T.–S. Balaban, A. T. Balaban, and D. Bonchev, A Topological Approach to the Predicting of Properties of Infinite Polymers. VI. Rational Formulas for the Normalized Wiener Index and a Comparison with Index J, *J. Mol. Structure* (*Theochem*) **2001**, *535*, 81–92.

[143] D. Bonchev, E. Markel, and A. Dekmezian, Topological Analysis of Long–Chain Branching Patterns in Polyolefins, *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1274–1285.

[144] D. Bonchev, E. Markel, and A. Dekmezian, Long–Chain Branch Polymer Dimensions: Application of Topology to the Zimm–Stockmayer Model, *Polymer* **2002**, *43*, 203–222.

[145] B. H. Zimm and W. H. Stockmayer, The Dimensions of Chain Molecules Containing Branches and Rings, *J. Chem. Phys.* **1949**, *17*, 1301–1314.

[146] P. E. Rouse, A Theory of the Linear Viscoelastic Properties of Dilute Solutions of Coiling Polymers, *J. Chem. Phys.* **1953**, *21*, 1272–1280; B. H. Zimm, Dynamics of Polymer Molecules in Dilute Solution. Viscoelasticity, Flow Birefringence and Dielectric Loss, *J. Chem. Phys.* **1956**, *24*, 269–278.

[147] B. E. Eichinger, Configurational Statistics of Gaussian Molecules, *Macromolecules* **1980**, *13*, 1–11.

[148] B. Mohar, The Laplacian Spectrum of Graphs: in: *Graph Theory*, *Combinatorics*, *and Applications*, Eds. Y. Alavi, G. Chartrand, O. R. Ollermann, and A. J. Schwenk, Wiley, 1991, pp. 871–898.

[149] Y. Yang, Graph Theory of Viscoelastic and Configurational Properties of Gaussian Chains, *Macromol. Theor. Simul.* **1998**, *7*, 521–549.

[150] D. J. Klein and M. Randić, Resistance Distance, *J. Math. Chem.* **1993**, *12*, 81–95.

[151] F. Markopoulou and L. Smolin, Causal evolution of spin networks, *arXiv*:gr–qc/ 9702025 v1, 13 Feb 1997.

[152] O. E. Polansky and D. Bonchev, The Wiener Number of Graphs. I. General Theory and Changes Due to Graph Operations, *Commun. Math. Comput. Chem* .(*MATCH*) **1986**, *21*, 133–186; II. Transfer Graphs and Some of Their Metric Properties, *ibid*. **1990**, *25*, 3–40.

[153] D. Bonchev, L. B. Kier, and O. Mekenyan, Self–Returning Walks and Fractional Electronic Charges of Atoms in Molecules, *Int. J. Quantum Chem.* **1993**, *46*, 635–649.

[154] D. Bonchev, W. A. Seitz, and E. Gordeeva, Relative Atomic Moments as Squared Principal Eigenvector Coefficients, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 237–242.

[155] I. Gutman, D. Bonchev, W. C. Seitz, and E. Gordeeva, Complementing the Proof of the Limit of Relative Atomic Moments, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 894–895.

[156] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Pathe Integrals*, McGraw Hill, New York, 1965.

[157] H. Nagao, K. Nishikawa, and S. Aono, The Hückel Model by Means of the Path Integral Method, *Chem. Phys. Lett.* **1992**, *190*, 97–103.

[158] D. Bonchev and E. Gordeeva, Topological Atomic Charges, Valencies and Bond Orders, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 383–395.

[159] D. Bonchev and L. B. Kier, Topological Atomic Indices and the Electron Charges in Alkanes, *J. Math. Chem.* **1992**, *9*, 75–85.

[160] D. Bonchev and W. C. Seitz, On Topological Charge Stabilization in Heterocyclic Compounds, *Russ. J. Heterocycl. Comp.* **1995**, *8*, 1011–1022.

[161] B. M. Gimarc and J. J. Ott; in: *Mathematics and Computational Concepts in Chemistry*, Ed. N. Trinajstić, Ellis Horwood, Chichester, U. K., 1986, pp. 74.

[162] A. Golbraikh, D. Bonchev, A. Tropsha, Novel Chirality Descriptors from Molecular Topology, *J. Chem. Inf. Comput. Sci.*, **2001**, *41*, 147–158.

[163] A. Golbraikh, D. Bonchev, and A. Tropsha, Novel ZE–Isomerism Descriptors Derived from Molecular Topology, *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 769–787.

[164] A. T. Balaban, O. Mekenyan and D. Bonchev, Unique Description of Chemical Structures Based on Hierarchically Ordered Extended Connectivities. I. Algorithms for Finding Graph Orbits and Cannonical Numbering of Atoms, Unique Description of Chemical Structures Based on Hierarchically Ordered Extended Connectivities (HOC Procedures), *J. Comput. Chem.* **1985**, *6*, 538–551.

[165] O. Mekenyan, D. Bonchev and A. T. Balaban, Unique Description of Chemical Structures Based on Hierarchically Ordered Extended Connectivities. II. Mathematical Proofs for the HOC Algorithm, Unique

Description of Chemical Structures Based on Hierchically Ordered Extended Connectivities (HOC Procedures), *J. Comput. Chem.* **1985**, *6*, 552–561.

[166] A. T. Balaban, O. Mekenyan, and D. Bonchev, Unique Description of Chemical Structures Based on Hierarchically Ordered Extended Connectivities. III. Topological Chemical, and Stereochemical Coding of Molecular Structure, and Unique Description of Chemical Structure Based on Hierarchically Ordered Extended Connectivities (HOC Procedures), *J. Comput. Chem.* **1985**, *6*, 562–569.

[167] D. Bonchev, O. Mekenyan, and A. T. Balaban, Unique Description of Chemical Structures Based on Hierarchically Ordered Extended Connectivities. IV. Recognition of Graph Isomorphism and Graph Symmetries, *Commun. Math. Comput. Chem* (*MATCH*) **1985**, *18*, 83–99.

[168] O. Mekenyan, D. Bonchev, and A. T. Balaban, Unique Description of Chemical Structures Based on Hierarchically Ordered Extended Connectivities. V. New Topological Indices, Ordering of Graphs, and Recognition of Graph Similarity, *J. Comput. Chem.* **1984**, *5*, 629–639.

[169] O. Mekenyan, D. Bonchev, and A. T. Balaban, Unique Description of Chemical Structures Based on Hierarchically Ordered Extended Connectivities. VI. Reflection in the 1H HMR Chemical Shifts of Condensed Benzenoid Hydrocarbons, *Chem. Phys. Lett.* **1984**, *109*, 85–88.

[170] O. Mekenyan, A. T. Balaban, and D. Bonchev, Unique Description of Chemical Structures Based on Hierarchically Ordered Extended Connectivities. VII. Condensed Benzenoid Hydrocarbons and Their 1H NMR Chemical Shifts, *J. Magn. Res.* **1985**, *63*, 1–13.

[171] H. L. Morgan, The Generation of a Unique Machine Description for Chemical Structures – A Technique Developed at Chemical Abstracts Service, *J. Chem. Doc.* **1965**, *5*, 107–113.

[172] D. Bonchev, O. Mekenyan, G. Protic, and N. Trinajstić, Application of Topological Indices to Gas Chromatographic Data: Calculation of the Retention Indices of Isomeric Alkylbenzenes, *J. Chromatogr.* **1979**, *176*, 149–156.

[173] D. Bonchev and O. Mekenyan, Comparability Graphs and Electronic Spectra of Condensed Benzenoid Hydrocarbons, *Chem. Phys. Lett.* **1983**, *98*, 134–138.

[174] D. Bonchev and O. Mekenyan, Comparability Graphs and Molecular Properties, *J. Chem. Soc.* Trans. Faraday II **1984**, *80*, 695–712.

[175] D. Bonchev, V. Kamenska, and O. Mekenyan, Comparability Graphs and Molecular Properties. III. C9– and C10 – Alkanes, *Int. J. Quantum Chem.* **1990**, *37*, 135–150.

[176] D. Bonchev, V. Kamenska, and O. Mekenyan, Comparability Graphs and Molecular Properties: IV. Generalizations and Applications, *J. Math. Chem.* **1990**, *5*, 43–72.

[177] O. Ivanciuc, private communication.

[178] S. Nikolić, G. Kovacević, A. Miličević, and N. Trinajstić, The Zagreb Indices 30 Years After, *Croat. Chem. Acta* **2003**, *76*, 113–124.

[179] A. Miličević and S. Nikolić, On Variable Zagreb Indices, *Croat. Chem. Acta* **2004**, *77*, 97–101.

[180] O. Mekenyan and D. Bonchev, On the Possible Mechanisms of Electronic Drug–Receptor Interaction, *Eur. J. Med. Chem.* **1989**, *24*, 397–408.

[181] O. Mekenyan, S. Karabunarliev, and D. Bonchev, The OASIS Concept for Predicting Biological Activity of Chemical Compounds, *J. Math. Chem.* **1990**, *4*, 207–215.

[182] O. Mekenyan, D. Bonchev, N. Trinajstić, and D. Peitchev. Modeling the Interaction of Small Organic Molecules with Biomacromolecules. II. A Generalized Concept for Biological Interactions, *Arzneim. Forsch.* **1986**, *36*, 421–424.

[183] O. Mekenyan, D. Peitchev, D. Bonchev, N. Trinajstić, and J. Dimitrova, III. Interaction of Benzoates with Anti–p–(p'–Azophenylazo) Benzoate Antibody, *Arzneim.l Forsch.* **1986**, *36*, 629–634.

[184] O. Mekenyan, D. Bonchev, D. H. Rouvray, D. Peitchev, and I. Bangov, IV. The In Vivo Interaction of 2– and 6–Substituted Purines With Murine Solid Tumor Adenocarcinoma CA 755, *Eur. J. Med. Chem.* **1991**, *26*, 305–312.

[185] O. Mekenyan, D. Bonchev and V. Enchev, V. Toxicity of Phenols to Alga "Lemna Minor", *Quant. Struct.– Act. Relat.* **1988**, *7*, 240–244.

[186] O. Mekenyan, C. Mercier, D. Bonchev, and J.–E. Dubois, Comparative Study of DARC/PELCO and OASIS Method. II. Modeling PNMT Inhibition Potency of Benzylamines and Amphetamines, *Eur. J. Med. Chem.* **1993**, *28*, 811–819.

[187] D. Bonchev, W. A. Seitz, C. F. Mountain, and A. T. Balaban, Modeling the Anticarcinogenic Action of Some Retinoid Compounds by Making Use of the OASIS Method. III. Inhibition of the Induction of Ornithine Decarboxylase by Arotinoids, *J. Med. Chem.* **1994**, *37*, 2300–2307.

[188] C. Hansch and A. Leo, *Substituent Constants for Correlation Analysis in Chemistry and Biology*, Wiley Interscience, New York, 1979.

[189] D. H. Rouvray and R. B. King (Eds.), *Topology in Chemistry. Discrete Mathematics of Molecules*, Horwood, Chichester, U. K., 2002 (Collection of reports on the First International Harry Wiener Conference held in Athens, Georgia, 1991).

[190] A. T. Balaban, Chemical Graphs. Part 35. Five New Topological Indices for the Branching of Tree–Like Graphs, *Theoret. Chim. Acta* (*Berlin*), **1979**, *5*, 239–261.

[191] A. T. Balaban, Highly Discriminating Distance–Based Topological Index *J*, *Chem. Phys. Lett.* **1982**, *89*, 399–404.

[192] W. T. Yee, K. Sakamoto, and Y. J. I'Haya, Information Theory of Molecular Properties. I. Theoretical Study of the Information Content of Organic Molecules, *Rep. Univ. Electro–Commun.* **1976**, *27*, 53–67.

[193] O. Mekenyan, Dynamic QSAR Techniques: Applications in Drug Design and Toxicology, Curr. Pharm. Design 2002, *8*, 1605–1621.194.

[194] O. G. Mekenyan, S. Dimitrov, P. Schmieder, and G. Veith, In Silico Modeling of Hazard Endpoints: Current Problems and Perspectives, *SAR and QSAR in Environ. Research* 2003,*14*, 361–371.

[195] O. N. Temkin and D. Bonchev, Classification and Coding of Chemical Reaction Mechanisms; in: *Mathematical Chemistry*, *Vol.* II, Eds. D. Bonchev and D. H. Rouvray, Gordon & Breach, London, U. K., 1992, pp. 53–97.

[196] O. N. Temkin, A. V. Zeigarnik, and D. Bonchev, Graph–Theoretical Models of Complex Reaction Mechanisms and Their Elementary Steps; in: *Graph–Theoretical Approaches to Chemical Reactivity*, Eds. D. Bonchev and O. Mekenyan, Kluwer Academic, Dordrecht, 1994, pp. 143–278.

[197] O. N. Temkin, A. V. Zeigarnik, and D. Bonchev, *Chemical Reaction Networks. A Graph Theoretical Approach.* CRC Press, Boca Raton, FL, 1996.

[198] D. Bonchev, O. N. Temkin, and D. Kamenski, On the Classification and Coding of Linear Reaction Mechanisms, *React. Kinet. Catal. Lett.* **1980**, *15*, 113–118.

[199] D. Bonchev, O. N. Temkin, and D. Kamenski, Graph Theoretical Classification and Coding of Chemical Reactions with a Linear Mechanism, *J. Comput. Chem.* **1982**, *3*, 95–111.

[200] D. Bonchev, D. Kamensky, and O. N. Temkin, Complexity Index for the Linear Mechanisms of Chemical Reactions, *J. Math. Chem.* **1987**, *1*, 345–388.

[201] O. N. Temkin, L. G. Brouk, and D. Bonchev, The Topological Structure of the Mechanisms of Complex Reactions, *Teor. Exper. Khim.* **1988**, *24*, 282–291.

[202] O. N. Temkin and D. Bonchev, Application of Graph Theory to Chemical Kinetics: Part 1. Kinetics of Complex Reactions, *J. Chem. Educ.* **1992**, *69*, 544–550.

[203] D. Kamenski, O. N. Temkin, and D. Bonchev, Reaction Network for the Epoxidation Reaction of Olefins With Organic Hydroperoxides, *Appl. Catal.* **1992**, *88*, 1–22.

[204] K. Gordeeva, D. Bonchev, D. Kamenski, and O. N. Temkin, Enumeration, Coding, and Complexity of Linear Reaction Mechanisms, *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 244–247.

[205] O. N. Temkin, A. V. Zeigarnik, and D. Bonchev, Applications of Graph Theory to Chemical Kinetics. Part 2. Topological Specificity of Single–Route Reaction Mechanisms, *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 729–737.

[206] A. V. Zeigarnik, O. N. Temkin, and D. Bonchev, Applications of Graph Theory to Chemical Kinetics. Part 3. Topological Specificity of Multiroute Reaction Mechanisms, *J. Chem. Inf. Comput. Sci.***1996**, *36*, 973–981.

[207] E. L. King and C. Altman, A Schematic Method of Deriving the Rate Laws for Enzyme–Catalyzed Reactions, *J. Phys. Chem.* **1956**, *60*, 1375–1378.

[208] M. I. Temkin, Graphical Method for Deriving of the Rate Laws of Complex Reactions, *Dokl. Acad. Nauk SSSR* **1965**, *165*, 615–618.

[209] J. Horiuti, Stoichiometrische Zahlen und die Kinetik der chemischen Reactionen, *J. Res. Inst. Catal. Hokkaido Univ.* **1957**, *5*, 1–26.