

Handbook of Chemoinformatics
Wiley–VCH
2003

Graph Theory in Chemistry

Ovidiu Ivanciuc

Sealy Center for Structural Biology and Molecular Biophysics
Department of Biochemistry and Molecular Biology
University of Texas Medical Branch
301 University Boulevard
Galveston, Texas 77555-0857, USA
Email: ovidiu_ivanciuc@yahoo.com
Email: iejmd@yahoo.com
URL: <http://ivanciuc.org/>
URL: <http://biochempress.com/>

O. Ivanciuc, Graph Theory in Chemistry. In: *Handbook of Chemoinformatics*, Ed.: J. Gasteiger, Wiley–VCH, 2003, pp. 103–138.

Ovidiu Ivanciuc studied organic chemistry at the Polytechnic University of Timisoara (Romania), where he started to investigate topological indices, structural descriptors, and QSAR with Professor Dan Ciubotariu. He continued investigations in these directions during the PhD studies with Professor Alexandru T. Balaban. After the Romanian Revolution against the communism (1989) he was appointed Assistant Professor and then Associate Professor of organic chemistry at the Polytechnic University of Bucharest. Between 1996 and 1999 he was visiting Professor at the University of Nice (France), collaborating with Professor Daniel Cabrol-Bass in various chemometrics projects. As a Welch Foundation fellow at the Texas A&M University (2000–2001) he performed quantum computations for very large systems with Professor Douglas J. Klein. Since 2001 he has investigated the structural determinants of protein allergenicity (<http://fermi.utmb.edu/SDAP>) in the department of computational biology and bioinformatics at the University of Texas Medical Branch, on Galveston island.

He is chief editor of the Internet Electronic Journal of Molecular Design (<http://www.biochempress.com>). A list of his publications and reprints can be downloaded from <http://www.ivanciuc.org>.



4 Graph Theory in Chemistry

Ovidiu Ivanciuc

4.1 Introduction

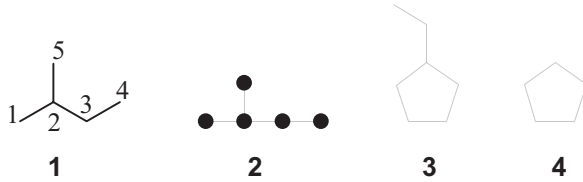
The wide applications of graph theory in physics, electronics, chemistry, biology, medicinal chemistry, economics, or the information sciences are mainly the effect of the seminal book of Harary [1]. A graph is usually represented in a graphical form as vertices interconnected by edges [1, 2]. Each graph vertex represents an object whereas the edge between two vertices represents the relationship between the two objects. In a chemical graph the objects can represent orbitals, atoms, bonds, groups of atoms, molecules, or collections of molecules. The edges of a chemical graph symbolize the interactions between chemical objects and are used to define chemical bonds, reactions, reaction mechanisms, kinetic models, or other relationships and transformations of the chemical objects. The rich literature on

chemical graphs and their applications [3–8] is an important guide in the exploration of the major applications of chemical graphs: topological indices and other structural indices for QSPR (quantitative structure–property relationships) and QSAR (quantitative structure–activity relationships) [9–16]; molecular orbital theory of conjugated compounds [17, 18]; structure of benzenoid hydrocarbons [19–22]; enumeration of isomers, constitutional symmetry perception and coding of chemical compounds [23–25]; kinetic and reaction graphs [26]; and computer-assisted synthesis design [27]. The main graph theory application in chemistry is the representation of chemical compounds as molecular graphs in which atoms are symbolized as vertices and the chemical bonds are represented as edges. Molecular graphs are used to represent chemical compounds and reaction mechanisms in a graphical form, to encode and search chemicals in databases, but their main application is in computing structural descriptors for drug design, virtual screening of chemical libraries, or for QSPR and QSAR models. Therefore, after a brief introduction to graph theory, we will present the main tools used in computing graph descriptors of molecular structure: parameters for molecules containing heteroatoms and multiple bonds, molecular matrixes, polynomials, spectra, and spectral moments. Other important applications of chemical graphs will be reviewed, such as enumeration of Kekulé structures, the topological resonance energy, and isomer enumeration. This contribution is a somewhat shortened and updated version of a chapter in the Encyclopedia of Computational Chemistry [28].

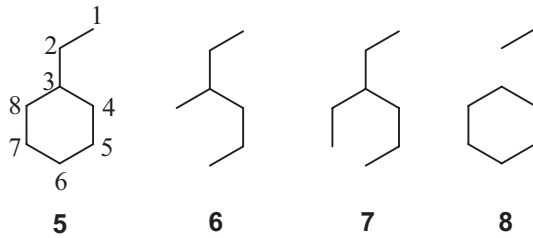
4.2

Elements of Graph Theory

A graph $G = G(V, E)$ is defined as an ordered pair consisting of two sets $V = V(G)$ and $E = E(G)$, where the elements of the set E define the binary relationship between the elements of the set V . In graphical form, the elements of the set V are represented as vertices whereas the elements of the set E are represented as edges connecting the vertices. The number of elements in $V(G)$, $N = |V(G)|$, defines the number of vertices N in the graph G and the number of elements in $E(G)$, $M = |E(G)|$, defines the number of edges M . The graph vertices are labeled from 1 to N , $V(G) = \{v_1, v_2, \dots, v_N\}$, and the edge connecting vertices v_i and v_j is denoted by e_{ij} . Two vertices v_i and v_j of a graph G are adjacent if there is an edge e_{ij} joining them; the vertices v_i and v_j are incident to the edge e_{ij} . Two distinct edges of G are adjacent if they have at least one vertex in common. As an example, consider graph **1**, that has five vertices and four edges, the vertex set $V(\mathbf{1}) = \{v_1, v_2, v_3, v_4, v_5\}$, and edge set $E(\mathbf{1}) = \{e_{1,2}, e_{2,3}, e_{3,4}, e_{2,5}\}$. Vertices v_2 and v_3 from graph **1** are adjacent and edges $e_{1,2}$, $e_{2,3}$, and $e_{2,5}$ are adjacent. The vertices are represented in a graphical form either as black circles or by the edge endpoints or connection with other edge(s). Usually, we will represent vertices as edge endpoints, as in graph **1**; however, in diagrams where the angle between edges is 180° the vertices must be represented as black circles, as in graph **2**, which is another graphical form of graph **1**.

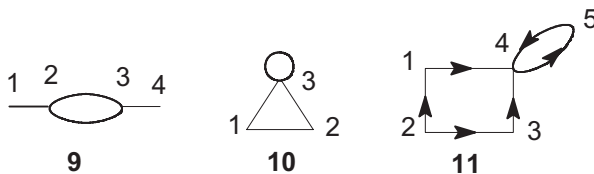


A subgraph H of a graph G is a graph whose vertices and edges are contained in G . If $V(H)$ is a subset of $V(G)$, $V(H) \subseteq V(G)$, and $E(H)$ is a subset of $E(G)$, $E(H) \subseteq E(G)$, then the subgraph $H = (V(H), E(H))$ is a subgraph of the graph $G = (V(G), E(G))$. The graph 4 is a subgraph of 3, because all vertices and edges of 4 are contained in 3. The subgraph $G - v_i$ is obtained by deleting from a graph G the vertex v_i and all its incident edges. The graph 6 is a subgraph of 5, obtained by deleting the vertex v_7 together with its incident edges $e_{6,7}$ and $e_{7,8}$. The subgraph $G - e_{ij}$ is obtained by deleting from graph G the edge e_{ij} . The graph 7 is a subgraph of 5, obtained by deleting the edge $e_{6,7}$, while the graph 8 is a subgraph of 5 obtained by deleting the edge $e_{2,3}$.

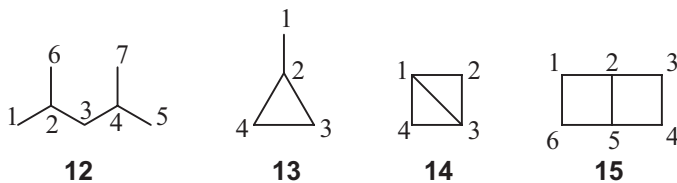


The degree of a vertex v_i , deg_i , is equal to the number of vertices adjacent to vertex v_i . The degree vector $Deg = Deg(G)$, in which the i th element represents the degree of the vertex v_i , collects the degrees for all vertices in a graph. The degree vector of graph 1 is $Deg(1) = \{1, 3, 2, 1, 1\}$ and the degree vector of graph 5 is $Deg(5) = \{1, 2, 3, 2, 2, 2, 2, 2\}$.

A multigraph contains pairs of vertices connected by more than one edge. A multiedge of multiplicity m is a set of m edges incident with the same pair of distinct vertices. The multigraph 9 has two vertices, v_2 and v_3 , connected by two edges. A loop is an edge joining a vertex with itself. The graph 10 contains a loop at vertex v_3 . A general graph may contain both multiple edges and loops, while a simple graph contains no loops or multiple edges. A digraph or directed graph $D = D(V, A)$ is defined by a set of nodes $V = V(D)$ and a set of arcs or directed edges $A = A(D)$, where the elements of the set A define the binary relation between the elements of the set V . A directed edge starting at node v_i and ending at node v_j is called an arc from v_i to v_j and is denoted by a_{ij} . Two arcs a_{ij} and a_{ji} are different. The digraph 11 has the node set $V(11) = \{v_1, v_2, v_3, v_4, v_5\}$ and the arc set $A(11) = \{a_{1,4}, a_{2,1}, a_{2,3}, a_{3,4}, a_{4,5}, a_{5,4}\}$.



A walk w in a graph G is defined as a sequence of vertices and edges $w = \{v_a, e_{ab}, v_b, e_{bc}, v_c, \dots, v_i, e_{ij}, v_j, \dots, v_m, e_{mn}, v_n\}$ beginning and ending with vertices, in which two consecutive vertices v_i and v_{i+1} are adjacent, and each edge e_{ij} is incident with the two vertices v_i and v_j preceding and following it, respectively; v_a is called the initial vertex of the walk and v_n is called the terminal vertex of the walk. Alternatively, a walk can also be defined as sequence of edges $w = \{e_{ab}, e_{bc}, \dots, e_{mn}\}$, in which two consecutive edges e_{ij} and e_{jk} are adjacent, or as a sequence of vertices $w = \{v_a, v_b, \dots, v_n\}$, in which two consecutive vertices v_i and v_{i+1} are adjacent. In a walk any edge of the graph can appear more than once. The length of a walk is the total number of edges that are contained in it. The following walks in the graph 12 are equivalent: $w_1(12) = \{v_1, e_{1,2}, v_2, e_{2,3}, v_3, e_{3,4}, v_4, e_{4,3}, v_3\}$; $w_2(12) = \{e_{1,2}, e_{2,3}, e_{3,4}, e_{4,3}\}$; $w_3(12) = \{v_1, v_2, v_3, v_4, v_3\}$. In these three identical walks of length four, $w_1(12) = w_2(12) = w_3(12)$, the initial vertex is v_1 , and the terminal vertex is v_3 .

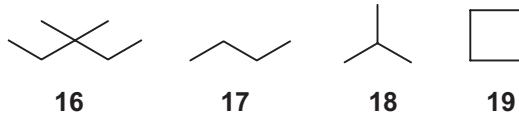


A closed walk or self-returning walk is a walk in which the initial and the terminal vertices coincide. A walk in which the initial and the terminal vertices are different is called an open walk or self-avoiding walk. In the graph 13 the walk $w_4(13) = \{v_1, v_2, v_4, v_3, v_4, v_3\}$ is open, while the walk $w_5(13) = \{v_4, v_2, v_3, v_4\}$ is closed. A path is a walk in which all vertices are distinct. The length of a path in a graph is equal to the number of edges along the path. A path of length 2 in the graph 14 is $w_6(14) = \{v_1, v_4, v_3\}$, while a path of length 3 in the same graph is $w_7(14) = \{v_2, v_3, v_1, v_4\}$. The graph 14 has no paths with length greater than 3. The length of a walk in a graph with two or more vertices has no limit.

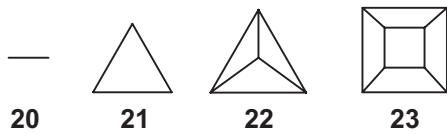
In a connected graph G every pair of vertices is joined by a path. In a disconnected graph G there is at least one pair of vertices $v_i, v_j \in V(G)$ with no path between them; these two vertices belong to different components of the graph G . The graph 8 is disconnected. The graph distance d_{ij} between a pair of vertices v_i and v_j from a connected graph G is defined as the length (number of edges) of the shortest path connecting the two vertices. The graph distance has the following properties: $d_{ii} = 0$ for all $v_i \in V(G)$; $d_{ij} > 0$ for all $v_i, v_j \in V(G)$; $d_{ij} = d_{ji}$ for all

$v_i, v_j \in V(G)$; $d_{ik} + d_{kj} \geq d_{ij}$ for all $v_i, v_j, v_k \in V(G)$. In the graph **15** there are three paths between vertices v_1 and v_6 : $w_8(\mathbf{15}) = \{v_1, v_6\}$ of length 1, $w_9(\mathbf{15}) = \{v_1, v_2, v_5, v_6\}$ of length 3, and $w_{10}(\mathbf{15}) = \{v_1, v_2, v_3, v_4, v_5, v_6\}$ of length 5. Because the length of the shortest path connecting vertices v_1 and v_6 is 1, the distance between the two vertices is 1, $d_{1,6} = d_{6,1} = 1$. The distance between two adjacent vertices is 1. The eccentricity $ecc(v_i)$ of a vertex v_i is defined as the maximum distance from the vertex v_i to any other vertex v_j in the graph G , $max(d_{ij})$ for all $v_j \in V(G)$. In the graph **15** the eccentricity of the vertex v_1 is 3, while the eccentricity of the vertex v_2 is 2. The diameter $diam(G)$ of a graph G is the maximum eccentricity. For the graphs **12**–**15**, $diam(\mathbf{12}) = 4$, $diam(\mathbf{13}) = 2$, $diam(\mathbf{14}) = 2$, and $diam(\mathbf{15}) = 3$.

A graph circuit or cycle is defined as a closed walk in which only the initial and terminal vertices coincide while all other vertices are distinct. In the graph **15** the walk $w_{11}(\mathbf{15}) = \{v_3, v_4, v_5, v_6, v_1, v_2, v_3\}$ is a cycle of length 6. The subgraph $G - C_i$ is obtained from the graph G by deleting all the vertices of the cycle C_i and their incident edges. The cyclomatic number μ is defined as the number of cycles in the graph, $\mu = M - N + 1$. For the graph **13** $\mu(\mathbf{13}) = 1$, for the graph **14** $\mu(\mathbf{14}) = 2$, for the graph **15** $\mu(\mathbf{15}) = 2$, while for the graph **12** $\mu(\mathbf{12}) = 0$. An acyclic graph, or a tree T , has $\mu = 0$. A k -tree is a tree with the maximum degree k . The graph **16** is a 4-tree, the graph **17** is a 2-tree, and the graph **18** is a 3-tree. A rooted tree is a tree in which one vertex (the root vertex) is distinct from the others.

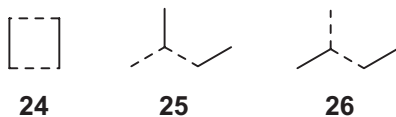


A chain or a linear graph L_N contains $N - 2$ vertices with degree 2 and two vertices with degree 1. The graph **17** is the linear graph L_4 . A star S_N is a tree with N vertices, $N - 1$ of them having degree 1. The graph **18** is the star graph S_4 . A spanning subgraph of a graph G is obtained from G by deleting one or more bonds. A subtree of a graph G is a subgraph of G which is a tree. A spanning tree of a graph G is a subtree of G that contains all vertices of G . The graph **13** is a spanning subgraph of the graph **14**, the graph **7** is a spanning tree of the graph **5**, while the graph **17** is a spanning tree of the graph **19**.



A graph in which every vertex has the degree k is a k -regular graph. The graph **20** is 1-regular, the graphs **4**, **19**, and **21** are 2-regular, and the graphs **22** and **23** are 3-regular. A complete graph K_N containing N vertices is a regular graph of degree

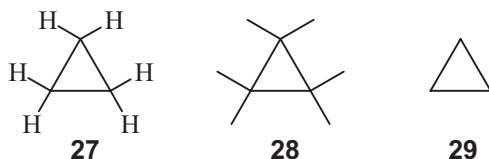
$N - 1$ with $N(N - 1)/2$ edges in which each vertex is adjacent to the remaining $N - 1$ vertices. The graph **20** is the complete graph K_2 , the graph **21** is K_3 , and the graph **22** is the complete graph K_4 . A graph with N vertices that all have the degree 2 defines a ring R_N . The graph **4** is the ring R_5 , the graph **19** is the ring R_4 , and the graph **21** represents R_3 .



A subgraph of a graph G consisting of k independent, mutually non-incident, edges represents a k -matching of G . The number of selections of k independent edges in G is denoted by $m(G, k)$; by definition, $m(G, 0) = 1$. The graph **24** is a 2-matching of the ring R_4 , and graphs **22** and **26** are two different 2-matchings of the graph **1**; the missing bonds are indicated with dashed lines. In the line graph $Li(G)$ of a graph G each vertex from $Li(G)$ corresponds to an edge from G , and two vertices from $Li(G)$ are adjacent if the corresponding edges from G are incident to a common vertex. The line graph of **1** is the graph **13**, $\mathbf{13} = Li(\mathbf{1})$, while the line graph of **13** is the graph **14**, $\mathbf{14} = Li(\mathbf{13})$. The line graph of a ring R_N is the same ring, $R_N = Li(R_N)$, while the line graph of a linear graph L_N is the linear graph with $N - 1$ vertices, $L_{N-1} = Li(L_N)$.

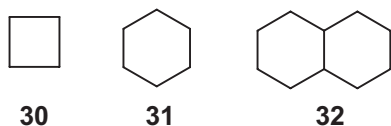
4.3 Molecular Graphs

Chemical compounds are usually represented as molecular graphs, i.e. non-directed, connected graphs in which vertices correspond to atoms and edges represent covalent bonds between atoms. The molecular graph model of the chemical structure emphasizes the chemical bonding pattern of atoms, whereas molecular geometry is neglected. Obviously, the molecular graph model is appropriate for prediction of physical, chemical, or biological properties that depend mainly on the bonding relationships between atoms. On the other hand, the molecular graph representation reflects mainly the connectivity of the atoms and is less suitable for the modeling of those properties that are determined mostly by the molecular geometry, conformation, or stereochemistry. The graph representation of chemical compounds must retain the features of the molecular structure that are relevant for the investigated physical, chemical, or biological property.

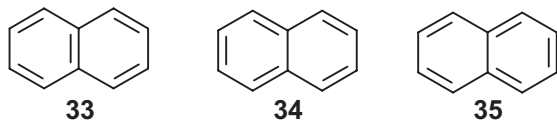


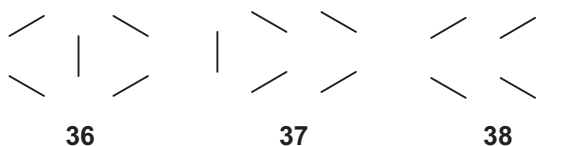
Cyclopropane can be represented, in different conventions, by graphs 27, 28, and 29. In the graph 28 hydrogen and carbon atoms are not differentiated, making this representation of little practical use. Also, the vertices representing the hydrogen atoms are redundant, because the structure of cyclopropane 27 can be reconstructed from its carbon skeleton 29. Therefore, the usual graph representation of an organic chemical compound is the hydrogen-depleted (or hydrogen-suppressed) molecular graph in which vertices correspond to non-hydrogen atoms and edges represent covalent bonds between non-hydrogen atoms. For hydrocarbons, the vertices in the molecular graph represent carbon atoms. Using this convention, alkanes are represented as 4-trees. The graph 1 is the hydrogen-suppressed molecular graph of 2-methylbutane, the graph 3 represents ethylcyclopentane, the graph 4 represents cyclopentane, the graph 12 represents 2,4-dimethylpentane, the graph 21 represents cyclopropane, the graph 22 represents tetrahedrane, and the graph 23 represents cubane.

Another important class of molecular graphs is the Hückel graph in which each vertex represents an sp^2 -hybridized carbon atom and each edge represents a conjugated carbon-carbon bond. The graph spectral theory has a close connection to the Hückel molecular orbital (HMO) theory, and the HMO study of conjugated compounds benefited from the application of graph theory [2, 4, 7, 17, 18]. Cyclobutadiene has the Hückel graph 30, benzene has the Hückel graph 31, and naphthalene has the Hückel graph 32.



A Kekulé structure of a conjugated hydrocarbon can be represented as a Kekulé graph. A Kekulé structure of a conjugated hydrocarbon is a structural formula, which may or may not include hydrogens, in which every carbon atom is sp^2 -hybridized and incident to exactly one double bond. A Kekulé graph is a disconnected graph consisting of K_2 components (isolated edges) representing the double bonds in the corresponding Kekulé structure. Naphthalene has three Kekulé structures, namely 33, 34, and 35, represented by the Kekulé graphs 36, 37, and 38. A 1-factor of a graph G is a spanning subgraph of G consisting of K_2 components. A Kekulé graph represents a 1-factor generated from the corresponding Kekulé structure.





The above examples of molecular graphs show clearly that the convention of representing chemical structures as molecular graphs is very important and has to be clearly specified in each study that uses molecular graphs. Depending on the convention used to translate atoms and bonds into vertices and edges, different chemical compounds can be represented by the same graph. When graph vertices represent sp^3 -hybridized carbon atoms and edges represent covalent carbon-carbon single bonds, graph 17 represents butane, graph 30 represents cyclobutane, graph 31 represents cyclohexane, and graph 32 represents decalin. On the other hand, when chemical compounds are represented as Hückel graphs, with each sp^2 -hybridized carbon atom represented as a graph vertex and each conjugated carbon-carbon bond represented as a graph edge, graph 17 represents butadiene, graph 30 represents cyclobutadiene, graph 31 represents benzene, and graph 32 represents naphthalene. The convention of representing molecules as graphs is also important for computing structural descriptors from molecular graphs. Although topological indices are usually computed from hydrogen-suppressed molecular graphs, some structural descriptors are derived from the whole, hydrogen-containing molecular graph. These examples demonstrated the importance of indicating in any model that uses chemical graphs the rules used to translate a chemical structure into a molecular graph.

4.4

Vertex- and Edge-Weighted Molecular Graphs

Although more than one thousand topological indices are used in QSPR, QSAR and design of chemical libraries, most are defined only for simple graphs that represent alkanes and cycloalkanes. The main reason for this limitation is that the mathematical theory of weighted graphs is not developed to a level high enough to ensure significant application to molecular graphs. QSAR studies, however, require the computation of topological indices for molecules containing heteroatoms and multiple bonds, which can be conveniently represented as vertex- and edge-weighted (VEW) molecular graphs. The computation of molecular matrixes and topological indices for VEW molecular graphs is usually performed with various sets of vertex and edge parameters [28, 29].

A VEW molecular graph $G = G(V, E, Sy, Bo, Vw, Ew, w)$ consists of a vertex set $V = V(G)$, an edge set $E = E(G)$, a set of chemical symbols for vertices $Sy = Sy(G)$, a set of topological bond orders for edges $Bo = Bo(G)$, a vertex weight set $Vw(w) = Vw(w, G)$, and an edge weight set $Ew(w) = Ew(w, G)$. The elements of

the vertex and edge weight sets are computed with the weighting scheme w . Usually, hydrogen atoms are not considered in the molecular graph, and in a VEW graph the weight of a vertex corresponding to a carbon atom is 0 and the weight of an edge corresponding to a carbon-carbon single bond is 1. Also, the topological bond order Bo_{ij} of an edge e_{ij} takes the value 1 for single bonds, 2 for double bonds, 3 for triple bonds and 1.5 for aromatic bonds. In a VEW graph G the length of a path p_{ij} between vertices v_i and v_j , $l(p_{ij}, w, G)$, is equal to the sum of the edge parameters $Ew(w)_{ij}$ for all edges along the path. The topological length of a path p_{ij} , $t(p_{ij}, G)$, in a VEW graph G is equal to the number of edges along the path. In a VEW graph, the distance $d(w)_{ij}$ between a pair of vertices v_i and v_j is equal to the length of the shortest path connecting the two vertices, $d(w)_{ij} = \min(l(p_{ij}, w))$. The topological distance td_{ij} between vertices v_i and v_j from a VEW graph G is equal to the minimum topological length of the paths connecting the two vertices, $td_{ij} = \min(t(p_{ij}))$, i.e. the minimum number of bonds between vertices v_i and v_j . In simple (non-weighted) graphs, the distance d_{ij} and topological distance td_{ij} are equal, while in weighted graphs usually this is not true.

In a weighting scheme w the vertex Vw and edge Ew parameters are computed from a property p_i associated with every vertex v_i from G , $v_i \in V(G)$, and the topological bond order Bo of all edges from the molecular graph. The vertex parameter $Vw(w)_i$ for the vertex v_i is [29]:

$$Vw(w)_i = 1 - p_C/p_i \quad (1)$$

and the edge parameter $Ew(w)_{ij}$ for the edge between vertices v_i and v_j is:

$$Ew(w)_{ij} = p_C p_C / Bo_{ij} p_i p_j \quad (2)$$

where p_i is the atomic property of vertex v_i , p_j is the atomic property of vertex v_j , and p_C is the atomic property for the carbon atom. Several weighting schemes for molecular graphs were defined by applying Eqs (1) and (2) to different atomic properties: Z , when p is the atomic number [28]; A , when p is the atomic mass [29]; P , when p is the atomic polarizability [29]; E , when p is the atomic electronegativity [29]; R , when p is the atomic radius [29]. In Table 4-1 we present a selected set of atomic properties used with different weighting schemes, while in Table 4-2 we give vertex parameters Vw computed with the Z , A , P , R , and E weighting schemes. The edge parameters Ew computed with the A and E weighting schemes are presented in Tables 4-3 and 4-4, respectively.

The AH weighting scheme uses the following equation to define the vertex parameter $Vw(AH)_i$ for the non-hydrogen atom i [29]:

$$Vw(AH)_i = 1 - A_C / (A_i + NoH_i A_H) = 1 - 12.011 / (A_i + 1.0079 NoH_i) \quad (3)$$

The edge parameter $Ew(AH)_{ij}$ for the bond between atoms i and j is defined by the equation:

Tab. 4-1 Selected set of atomic properties used with different weighting schemes: atomic number Z , atomic mass A , polarizability α (\AA^3), atomic radius r (\AA), and electronegativity χ [29]

Atom	Z	A	α	r	χ
B	5	10.811	3.03	1.45	2.02
C	6	12.011	1.76	1.21	2.55
N	7	14.007	1.10	1.03	3.12
O	8	15.999	0.802	0.93	3.62
F	9	18.998	0.557	0.82	4.23
Si	14	28.086	5.38	1.75	1.87
P	15	30.974	3.63	1.54	2.22
S	16	32.066	2.90	1.43	2.49
Cl	17	35.453	2.18	1.30	2.82
As	33	74.922	4.31	1.63	2.11
Se	34	78.960	3.77	1.56	2.31
Br	35	79.904	3.05	1.45	2.56
Te	52	127.60	5.5	1.77	2.08
I	53	126.90	4.7	1.68	2.27

Tab. 4-2 Vertex parameters Vw computed with the Z , A , P , R , and E weighting schemes

Atom	$Vw(Z)$	$Vw(A)$	$Vw(P)$	$Vw(R)$	$Vw(E)$
B	-0.200	-0.111	0.419	0.166	-0.262
C	0.000	0.000	0.000	0.000	0.000
N	0.143	0.143	-0.600	-0.175	0.183
O	0.250	0.249	-1.195	-0.301	0.296
F	0.333	0.368	-2.160	-0.476	0.397
Si	0.571	0.572	0.673	0.309	-0.364
P	0.600	0.612	0.515	0.214	-0.149
S	0.625	0.625	0.393	0.154	-0.024
Cl	0.647	0.661	0.193	0.069	0.096
As	0.818	0.840	0.592	0.258	-0.209
Se	0.824	0.848	0.533	0.224	-0.104
Br	0.829	0.850	0.423	0.166	0.004
Te	0.885	0.906	0.680	0.316	-0.226
I	0.887	0.905	0.626	0.280	-0.123

$$\begin{aligned}
 Ew(AH)_{ij} &= A_C A_C / B_{0ij} (A_i + NoH_i A_H) (A_j + NoH_j A_H) \\
 &= 12.011 \quad 12.011 / B_{0ij} (A_i + 1.0079 NoH_i) (A_j + 1.0079 NoH_j) \quad (4)
 \end{aligned}$$

where $A_C = 12.011$ is the atomic mass for carbon, $A_H = 1.0079$ is the atomic mass for hydrogen, NoH_i is the number of hydrogen atoms bonded to the non-hydrogen atom i , and NoH_j is the number of hydrogen atoms bonded to the heavy atom j .

Tab. 4-3 Edge parameters $Ew(A)$ computed with the atomic mass weighting scheme A

<i>Atom_i</i>	<i>Atom_j</i>	<i>Single</i>	<i>Double</i>	<i>Triple</i>	<i>Aromatic</i>
C	C	1.000	0.500	0.333	0.667
C	N	0.857	0.429	0.286	0.572
C	O	0.751	0.375		
C	F	0.632			
C	Si	0.428			
C	P	0.388			
C	S	0.375	0.187		0.250
C	Cl	0.339			
C	Se	0.152			
C	Br	0.150			
C	Te	0.094			
C	I	0.095			
N	N	0.735	0.368		0.490
N	O	0.644	0.322		0.429
O	S	0.281	0.141		

Tab. 4-4 Edge parameters $Ew(E)$ computed with the atomic electronegativity weighting scheme E

<i>Atom_i</i>	<i>Atom_j</i>	<i>Single</i>	<i>Double</i>	<i>Triple</i>	<i>Aromatic</i>
C	C	1.000	0.500	0.333	0.667
C	N	0.817	0.409	0.272	0.545
C	O	0.704	0.352		
C	F	0.603			
C	Si	1.364			
C	P	1.149			
C	S	1.024	0.512		0.683
C	Cl	0.904			
C	Se	1.104			
C	Br	0.996			
C	Te	1.226			
C	I	1.123			
N	N	0.668	0.334		0.445
N	O	0.576	0.288		0.384
O	S	0.721	0.361		

4.5

Molecular Graph Matrixes

Molecular graphs are widely used to represent the chemical structure of covalent compounds in a graphical form, and this convention is widely used in chemistry textbooks and research papers. The molecular graph is, however, a non-numerical

representation of the chemical structure, and the computation of topological indices for QSAR requires a numerical description of graphs. Graphs can be represented in algebraic form as matrixes [7, 15, 29]. This numerical description of the structure of chemical compounds is essential for computer manipulation of molecules and for calculation of various topological indices.

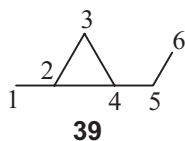
4.5.1

The Adjacency Matrix

The adjacency matrix $A = A(G)$ of a graph G with N vertices is the square $N \times N$ symmetric matrix whose elements $[A]_{ij}$ are defined as:

$$[A(G)]_{ij} = \begin{cases} 1 & \text{if } i \neq j \text{ and } e_{ij} \in E(G) \\ 0 & \text{if } i = j \text{ or } e_{ij} \notin E(G) \end{cases} \quad (5)$$

where $E(G)$ represents the set of edges of G . In the adjacency matrix $A(G)$ the row i and column i correspond to vertex v_i from G . As an example, the molecular graph and the adjacency matrix of 1-ethyl-2-methylcyclopropane **39** are given. The graph vertices are labeled from 1 to 6, and each vertex v_i is represented in the matrix $A(39)$ by the row i and column i , respectively.

 $A(39)$

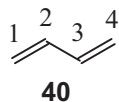
	1	2	3	4	5	6
1	0	1	0	0	0	0
2	1	0	1	1	0	0
3	0	1	0	1	0	0
4	0	1	1	0	1	0
5	0	0	0	1	0	1
6	0	0	0	0	1	0

The adjacency matrix $A(w, G)$ of a vertex- and edge-weighted molecular graph G with N vertices is the square $N \times N$ real symmetric matrix whose element $[A(w)]_{ij}$ is [29]:

$$[A(w, G)]_{ij} = \begin{cases} Vw(w)_i & \text{if } i = j \\ Ew(w)_{ij} & \text{if } e_{ij} \in E(G) \\ 0 & \text{if } e_{ij} \notin E(G) \end{cases} \quad (6)$$

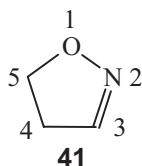
where $Vw(w)_i$ is the weight of the vertex v_i , $Ew(w)_{ij}$ is the weight of the edge e_{ij} , and w is the weighting scheme used to compute Vw and Ew . The first example of a

VEW adjacency matrix is for butadiene **40**; in the corresponding matrix $A(40)$ the weight for the double bonds is 0.5.



$A(40)$

	1	2	3	4
1	0.0	0.5	0.0	0.0
2	0.5	0.0	1.0	0.0
3	0.0	1.0	0.0	0.5
4	0.0	0.0	0.5	0.0



The adjacency matrix of 2-isoxazoline **41** computed with the electronegativity weighting scheme E is:

$A(E, 41)$

	1	2	3	4	5
1	0.296	0.576	0.000	0.000	0.704
2	0.576	0.183	0.409	0.000	0.000
3	0.000	0.409	0.000	1.000	0.000
4	0.000	0.000	1.000	0.000	1.000
5	0.704	0.000	0.000	1.000	0.000

4.5.2

The Burden Matrix

The molecular matrix originally proposed by Burden for computation of graph spectra (matrix eigenvalues) is a modified adjacency matrix obtained from the hydrogen-suppressed molecular graph of an organic compound [30]. The BCUT descriptors [31], derived from the graph spectra of the Burden matrix B , are extensively used in combinatorial chemistry, virtual screening, diversity measurement, and QSAR. An extension of the Burden matrix was defined by inserting on the main diagonal of the matrix a vertex structural descriptor VSD , representing a vector of experimental or computed atomic properties [32]. The rules defining the Burden matrix $B(VSD) = B(VSD, G)$ of a graph G with N vertices are:

1. The diagonal elements of B, $[B]_{ii}$, are computed with the formula:

$$[B(VSD)]_{ii} = VSD_i \quad (7)$$

where VSD_i is a vertex structural descriptor of the vertex v_i that reflects the local structure of the corresponding atom i .

2. The non-diagonal element $[B]_{ij}$, representing an edge e_{ij} connecting vertices v_i and v_j , has the value 0.1 for single bonds, 0.2 for a double bond, 0.3 for a triple bond, and 0.15 for an aromatic delocalized bond.
3. The value of a non-diagonal element $[B]_{ij}$ representing an edge e_{ij} connecting vertices v_i and v_j is augmented by 0.01 if either vertex v_i or vertex v_j have the degree 1.
4. All other non-diagonal elements $[B]_{ij}$ are set equal to 0.001 (these elements are set to 0 in the adjacency matrix A and correspond to pairs of nonbonded vertices in the molecular graph G).

4.5.3

The Laplacian Matrix

For a graph G we define the diagonal matrix, $DEG = DEG(G)$, whose i th entry is equal to the degree of the vertex v_i , deg_i , and all other elements are equal to zero. The Laplacian matrix of a simple graph G , $L = L(G)$, is a symmetric matrix defined with the equation [33, 34]:

$$L(G) = DEG(G) - A(G) \quad (8)$$

where $A(G)$ is the adjacency matrix of the graph G . The elements of the Laplacian matrix are:

$$[L]_{ij} = \begin{cases} \text{deg}_i & \text{if } i = j \\ -1 & \text{if } e_{ij} \in E(G) \\ 0 & \text{if } e_{ij} \notin E(G) \end{cases} \quad (9)$$

The Laplacian matrix of 1-ethyl-2-methylcyclopropane **39** is:

L(39)

	1	2	3	4	5	6
1	1	-1	-0	-0	-0	-0
2	-1	3	-1	-1	-0	-0
3	-0	-1	2	-1	-0	-0
4	-0	-1	-1	3	-1	-0
5	-0	-0	-0	-1	2	-1
6	-0	-0	-0	-0	-1	1

The above formula for L can be applied only to simple molecular graphs. The definition of the Laplacian matrix for VEW graphs uses the vertex valence instead of deg . The valence of the vertex v_i , $val(w)_i = val(w, G)_i$, is defined as the sum of the weights $Ew(w)_{ij}$ of all edges e_{ij} incident with vertex v_i [34]:

$$\mathbf{val}(w)_i = \sum_{e_{ij} \in E(G)} Ew(w)_{ij} \quad (10)$$

where w is the weighting scheme used to compute the Ew parameters. Alternatively, the valence of the vertex v_i may be computed as the sum of the non-diagonal elements in the row i , or column i , from the adjacency matrix $A(w) = A(w, G)$, of a molecular graph G with N vertices [34]:

$$\mathbf{val}(w)_i = \sum_{\substack{j=1 \\ j \neq i}}^N [A(w)]_{ij} = \sum_{\substack{j=1 \\ j \neq i}}^N [A(w)]_{ji} \quad (11)$$

The set of valence values for all vertices in a graph forms the vector $Val(w) = Val(w, G)$ whose i th element represents the valence of the vertex v_i . In alkanes and cycloalkanes the degree of a vertex v_i , deg_i , is identical with the valence of that vertex, val_i , while for molecules containing heteroatoms and multiple bonds, represented as vertex- and edge-weighted molecular graphs, this equality is not true.

The Laplacian matrix $L(w) = L(w, G)$ of a vertex- and edge-weighted molecular graph G with N vertices is the square $N \times N$ real symmetric matrix whose element $[L(w)]_{ij}$ is defined as [34]:

$$[L(w)]_{ij} = \begin{cases} \mathbf{val}(w)_i & \text{if } i = j \\ -Ew(w)_{ij} & \text{if } e_{ij} \in E(G) \\ 0 & \text{if } e_{ij} \notin E(G) \end{cases} \quad (12)$$

where $Ew(w)_{ij}$ is the weight of the edge e_{ij} , and w is the weighting scheme used to compute the parameters Ew .

4.5.4

The Distance Matrix

The distance matrix $D = D(G)$ of a graph G with N vertices is the square $N \times N$ symmetric matrix whose elements are defined as [7, 15]:

$$[D]_{ij} = \begin{cases} d_{ij} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \quad (13)$$

where d_{ij} is the graph distance (number of edges on the shortest path) between vertices v_i and v_j . As an example, the distance matrix of the graph 39, $D(39)$, is shown:

D(39)

	1	2	3	4	5	6
1	0	1	2	2	3	4
2	1	0	1	1	2	3
3	2	1	0	1	2	3
4	2	1	1	0	1	2
5	3	2	2	1	0	1
6	4	3	3	2	1	0

The distance matrix $D(w) = D(w, G)$ of a vertex- and edge-weighted molecular graph G with N vertices is the symmetric square $N \times N$ matrix with real elements defined with the formula [35]:

$$[\mathbf{D}(w, G)]_{ij} = \begin{cases} d(w)_{ij} & \text{if } i \neq j \\ Vw(w)_i & \text{if } i = j \end{cases} \quad (14)$$

where $d(w)_{ij}$ is the distance between vertices v_i and v_j , $Vw(w)_i$ is the weight of the vertex v_i , and w is the weighting scheme used to compute the parameters Vw and Ew .

4.5.5

The Reciprocal Distance Matrix

The reciprocal distance matrix $RD(w) = RD(w, G)$ of a vertex- and edge-weighted molecular graph G with N vertices is the square $N \times N$ symmetric matrix with real elements defined with the equation [29, 36]:

$$[\mathbf{RD}(w, G)]_{ij} = \begin{cases} 1/[\mathbf{D}(w, G)]_{ij} & \text{if } i \neq j \\ [\mathbf{D}(w, G)]_{ii} & \text{if } i = j \end{cases} \quad (15)$$

where $[D(w)]_{ij}$ is the graph distance between vertices v_i and v_j , $[D(w)]_{ii}$ is the diagonal element corresponding to vertex v_i , and w is the weighting scheme used to compute the parameters Vw and Ew . The reciprocal distance matrix of the graph **39** corresponding to 1-ethyl-2-methylcyclopropane is:

D(39)

	1	2	3	4	5	6
1	0	1	1/2	1/2	1/3	1/4
2	1	0	1	1	1/2	1/3
3	1/2	1	0	1	1/2	1/3
4	1/2	1	1	0	1	1/2
5	1/3	1/2	1/2	1	0	1
6	1/4	1/3	1/3	1/2	1	0

4.5.6

The Detour Matrix

The detour matrix (or the maximum path matrix MP) of a graph G with N vertices is the square $N \times N$ symmetric matrix whose element $[\Delta]_{ij}$ is defined as [37]:

$$[\Delta]_{ij} = \begin{cases} \max(l(p_{ij})) & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \quad (16)$$

where $l(p_{ij})$ is the length of the path p_{ij} , and $\max(l(p_{ij}))$ is the length of the longest path connecting the vertices v_i and v_j . For trees the detour matrix is identical with the distance matrix. While for the computation of the distance matrix there are many efficient algorithms, the detour matrix is difficult to determine, especially for polycyclic graphs.

4.5.7

The Detour-Distance Matrix

The detour-distance matrix $\Delta - D$ (originally called the maximum/minimum path matrix MmP) of a graph G with N vertices is the square $N \times N$ non-symmetric matrix that collects in its upper triangle the elements of the detour matrix while the lower triangle elements are identical to those in the distance matrix [37]:

$$[\Delta - D]_{ij} = \begin{cases} [\Delta]_{ij} & \text{if } i < j \\ [D]_{ij} & \text{if } i > j \\ 0 & \text{if } i = j \end{cases} \quad (17)$$

4.5.8

The Distance-Detour Quotient Matrix

Another mixing of the detour and distance matrixes into a single matrix results in the distance-detour quotient matrix D/Δ whose elements are defined with the formula [38]:

$$[D/\Delta]_{ij} = \begin{cases} [D]_{ij}/[\Delta]_{ij} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \quad (18)$$

4.5.9

The Distance-Valence Matrix

The distance-valence matrix of a vertex- and edge-weighted graph G with N vertices, $Dval(p, q, r, w) = Dval(p, q, r, w, G)$, is a square $N \times N$ matrix, whose entries $[Dval(p, q, r, w)]_{ij}$ are equal to [39]:

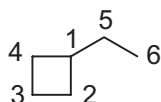
$$[Dval(p, q, r, w, G)]_{ij} = \begin{cases} d(w)_{ij}^p \text{val}(w)_i^q \text{val}(w)_j^r & \text{if } i \neq j \\ Vw(w)_i \text{val}(w)_i^{q+r} & \text{if } i = j \end{cases} \quad (19)$$

where $Vw(w)_i$ is the weight of vertex v_i , $d(w)_{ij}$ is the graph distance between vertices v_i and v_j , $val(w)_i$ is the valence of vertex v_i , all computed with the weighting scheme w . For vertex- and edge-weighted molecular graph the definition of the $Dval$ matrix was formulated in analogy with that for the reciprocal distance matrix RD , in such a way that $Dval(-1, 0, 0, w, G) = RD(w, G)$. In the particular case when $p = 1$ and $q = r = 0$ the $Dval(p, q, r, w)$ matrix is identical with the distance matrix $D(w)$. From the definition of the $Dval$ matrix one can see that non-symmetric matrixes can be obtained if $q \neq r$.

4.5.10

The Resistance Distance Matrix

Applying results from the electrical network theory, Klein and Randić introduced a new distance function on graphs called the resistance distance [40]. This novel graph distance was utilized to define the resistance-distance matrix Ω , proposed as an alternative to the distance matrix D . For the computation of the molecular matrix Ω , Klein and Randić superposed onto the molecular graph G an electrical network of resistors, in such a way that carbon atoms become nodes in the network and carbon-carbon single bonds are represented as 1 ohm resistors; the matrix element Ω_{ij} is equal to the effective electrical resistance between the vertices v_i and v_j . From the theory of electrical networks it is easy to determine that in the case of acyclic hydrocarbons (i.e. alkanes, alkenes, alkynes, etc.), the resistance-distance matrix Ω is identical with the distance matrix D , while in the case of cyclic compounds the two matrixes are different.

**42**

The computation of the resistance distance matrix Ω of a simple (non-weighted) graph is presented for ethylcyclobutane **42**, giving:

$\Omega(1)$

	1	2	3	4	5	6
1	0	3/4	1	3/4	1	2
2	3/4	0	3/4	1	7/4	11/4
3	1	3/4	0	3/4	2	3
4	3/4	1	3/4	0	7/4	11/4
5	1	7/4	2	7/4	0	1
6	2	11/4	3	11/4	1	0

We present here a more general definition and procedure for the computation of the resistance-distance matrix Ω of weighted molecular graphs, corresponding to

organic compounds with heteroatoms and multiple bonds [41]. Consider an electrical network of resistors in which a node (vertex) v_i corresponds to a vertex (with the same label) in the molecular graph G , while each chemical bond e_{ij} from the molecular graph is represented as a resistor between nodes v_i and v_j . Each resistor has a value $Ew_{ij}(w)$ (in ohm) depending on the chemical nature of the atoms represented by vertices v_i and v_j , and on the type of the chemical bond between them; the $Ew_{ij}(w)$ parameter is computed using the weighting schemes w proposed in the literature. The resistance distance matrix $\Omega(w) = \Omega(w, G)$ of a graph G with N vertices is the square $N \times N$ symmetric matrix whose non-diagonal element Ω_{ij} is equal to the electrical resistance $r(w)_{ij}$ between vertexes v_i and v_j :

$$[\Omega(w)]_{ij} = \begin{cases} r(w)_{ij} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \quad (20)$$

where $r(w)_{ij}$ is measured in ohm. The algorithm for computing the resistance-distance matrix of a vertex- and edge-weighted molecular graph comprises the following steps:

1. Set up the edge-weighted adjacency matrix of the molecular graph G that contains heteroatoms and multiple bonds:

$$\mathbf{A}_{ij}(w, G) = \begin{cases} 0 & \text{if } i = j \\ Ew_{ij}(w) & \text{if } e_{ij} \in E(G) \\ 0 & \text{if } e_{ij} \notin E(G) \end{cases} \quad (21)$$

2. Compute the reverse adjacency matrix $\mathbf{A}^-(w, G)$:

$$\mathbf{A}_{ij}^-(w, G) = \begin{cases} 0 & \text{if } i = j \\ 1/\mathbf{A}_{ij}(w, G) & \text{if } e_{ij} \in E(G) \\ 0 & \text{if } e_{ij} \notin E(G) \end{cases} \quad (22)$$

3. Obtain the Laplacian matrix of the reverse adjacency matrix:

$$\mathbf{L}_{ij}(w, G) = \begin{cases} \sum_{\{i,k\} \in E(G)} \mathbf{A}_{ik}^-(w, G) & \text{if } i = j \\ -\mathbf{A}_{ij}^-(w, G) & \text{if } e_{ij} \in E(G) \\ 0 & \text{if } e_{ij} \notin E(G) \end{cases} \quad (23)$$

4. Calculate the eigenvalues and eigenvectors of the weighted Laplacian matrix $L(w, G)$ of the molecular graph G with N vertices:

$$\mathbf{L} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^t \quad (24)$$

where U is an $N \times N$ column matrix of eigenvectors of the weighted Laplacian matrix L , U^t is the transpose matrix, and Λ is an $N \times N$ diagonal matrix con-

taining on the main diagonal the eigenvalues of L ; the eigenvalue $[\Lambda]_{ii}$ corresponds to the eigenvector from the i th column of matrix U . For any connected molecular graph the Laplacian matrix L has all eigenvalues positive except for one which is 0.

5. The $N \times N$ diagonal matrix V is computed from the eigenvalues of L :

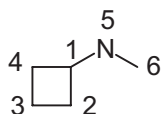
$$[V]_{ij} = \begin{cases} 0 & \text{if } [\Lambda]_{ij} = 0 \\ [\Lambda]_{ij}^{-1} & \text{if } [\Lambda]_{ij} \neq 0 \end{cases} \quad (25)$$

6. The generalized inverse of L is the matrix Γ which is 0 on its null eigenspace and the “true” inverse on the subspace orthogonal to this null space:

$$\Gamma = UVU^t \quad (26)$$

7. The resistance-distance matrix is obtained from Γ :

$$\Omega_{ij}(w, G) = \begin{cases} \Gamma_{ii}(w, G) - 2\Gamma_{ij}(w, G) + \Gamma_{jj}(w, G) & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \quad (27)$$



43

The computation of the resistance distance matrix for cyclobutylmethylamine **43**, using edge parameters $Ew(P)$ computed with the molecular polarizability weighting scheme P , with a parameter for carbon–nitrogen bonds $Ew(C-N, P) = 1.60$, gives:

$\Omega(P, 43)$

	1	2	3	4	5	6
1	0	0.75	1.00	0.75	1.60	3.20
2	0.75	0	0.75	1.00	2.35	3.95
3	1.00	0.75	0	0.75	2.60	4.20
4	0.75	1.00	0.75	0	2.35	3.95
5	1.60	2.35	2.60	2.35	0	1.60
6	3.20	3.95	4.20	3.95	1.60	0

4.5.11

The Electrical Conductance Matrix

The reciprocal of the resistance distance matrix Ω , with non-diagonal elements equal to $1/\Omega_{ij}$, defines the electrical conductance matrix $EC(w) = EC(w, G)$ [41]:

$$[\mathbf{EC}(w)]_{ij} = \begin{cases} 1/r(w)_{ij} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \quad (28)$$

where w is the weighting scheme used to compute the parameters $Ew(w)_{ij}$ representing the value of the resistor between vertices v_i and v_j . The electrical conductance matrix EC of ethylcyclobutane **42** is:

EC(42)

	1	2	3	4	5	6
1	0	4/3	1	4/3	1	1/2
2	4/3	0	4/3	1	4/7	4/11
3	1	4/3	0	4/3	1/2	1/3
4	4/3	1	4/3	0	4/7	4/11
5	1	4/7	1/2	4/7	0	1
6	1/2	4/11	1/3	4/11	1	0

4.5.12

The Distance-Path Matrix

Consider the vertex- and edge-weighted graph G with N vertices and its distance matrix $D(w) = D(w, G)$ computed with the weighting scheme w . The distance-path matrix of the weighted graph G , $D_p(w) = D_p(w, G)$, is the square $N \times N$ symmetric matrix whose element $[D_p(w)]_{ij}$ is defined with the formula [42]:

$$[D_p(w, G)]_{ij} = [D(w, G)]_{ij}([D(w, G)]_{ij} + 1)/2 \quad (29)$$

4.5.13

The Reciprocal Distance-Path Matrix

The reciprocal distance-path matrix of a VEW graph G with N vertices, $RD_p(w) = RD_p(w, G)$, is the square $N \times N$ symmetric matrix whose element $[RD_p(w)]_{ij}$ is equal to the reciprocal of the corresponding distance-path matrix element, $1/[D_p(w)]_{ij}$, for non-diagonal elements, and is equal to $[D_p(w)]_{ii}$ for the diagonal elements [42]:

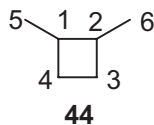
$$[\mathbf{RD}_p(w, G)]_{ij} = \begin{cases} 1/[\mathbf{D}_p(w, G)]_{ij} & \text{if } i \neq j \\ [\mathbf{D}_p(w, G)]_{ii} & \text{if } i = j \end{cases} \quad (30)$$

4.5.14

The Distance Complement Matrix

The distance complement matrix $DC(w, G)$ of a VEW graph G with N vertices is the square $N \times N$ symmetric matrix whose elements are defined as [43]:

$$[\mathbf{DC}(w, G)]_{ij} = \begin{cases} N - [\mathbf{D}(w, G)]_{ij} & \text{if } i \neq j \\ Vw(w)_i & \text{if } i = j \end{cases} \quad (31)$$



The distance complement matrix DC of 1,2-dimethylcyclobutane **44** is:

$DC(44)$

	1	2	3	4	5	6
1	0	5	4	5	5	4
2	5	0	5	4	4	5
3	4	5	0	5	3	4
4	5	4	5	0	4	3
5	5	4	3	4	0	3
6	4	5	4	3	3	0

4.5.15

The Reciprocal Distance Complement Matrix

Using a procedure similar to that employed in the computation of reciprocal matrixes one can transform the distance complement matrix into the reciprocal distance complement matrix. The reciprocal distance complement matrix $RDC(w) = RDC(w, G)$ of a vertex- and edge-weighted molecular graph G with N vertices is the square $N \times N$ symmetric matrix with real elements defined with the equation [35, 44]:

$$[\mathbf{RDC}(w, G)]_{ij} = \begin{cases} 1/[\mathbf{DC}(w, G)]_{ij} & \text{if } i \neq j \\ [\mathbf{DC}(w, G)]_{ii} & \text{if } i = j \end{cases} \quad (32)$$

where $[DC(w)]_{ij}$ is the graph distance complement between vertices v_i and v_j , $[DC(w)]_{ii}$ is the diagonal element corresponding to vertex v_i , and w is the weighting scheme used to compute the vertex and edge parameters Vw and EW .

4.5.16

The Complementary Distance Matrix

The complementary distance matrix CD is another matrix in which the value of the matrix elements corresponding to pairs of vertices decreases when the distance between the vertices increases. The complementary distance matrix $CD(w) = CD(w, G)$ of a vertex- and edge-weighted molecular graph G with N vertices is the

square $N \times N$ symmetric matrix whose elements are defined as [35, 45]:

$$[\mathbf{CD}(w, G)]_{ij} = \begin{cases} d(w)_{\max} + d(w)_{\min} - [\mathbf{D}(w, G)]_{ij} & \text{if } i \neq j \\ Vw(w)_i & \text{if } i = j \end{cases} \quad (33)$$

where $[D(w)]_{ij}$ is the ij th element of the distance matrix $D(w)$ which is equal to the graph distance between vertices v_i and v_j , $d(w)_{\max}$ is the maximum distance between two distinct graph vertices (the graph diameter), and $d(w)_{\min}$ is the minimum distance between two distinct graph vertices (equal to 1 for alkanes and cycloalkanes):

$$d(w)_{\max} = \max\{d(w)_{ij}, v_i, v_j \in V(G), v_i \neq v_j\} \quad (34)$$

$$d(w)_{\min} = \min\{d(w)_{ij}, v_i, v_j \in V(G), v_i \neq v_j\} \quad (35)$$

4.5.17

The Reciprocal Complementary Distance Matrix

The reciprocal complementary distance matrix $RCD(w, G)$ of a vertex- and edge-weighted molecular graph G with N vertices is the square $N \times N$ symmetric matrix defined with the equation [35, 45]:

$$[\mathbf{RCD}(w, G)]_{ij} = \begin{cases} 1/[\mathbf{CD}(w, G)]_{ij} & \text{if } i \neq j \\ [\mathbf{CD}(w, G)]_{ii} & \text{if } i = j \end{cases} \quad (36)$$

4.5.18

The Reverse Wiener Matrix

The reverse Wiener matrix $RW(w, G)$ of a vertex- and edge-weighted molecular graph G with N vertices is the square $N \times N$ symmetric matrix whose elements are obtained by subtracting from d_{\max} each d_{ij} value in the distance matrix [46]:

$$[\mathbf{RW}(w, G)]_{ij} = \begin{cases} d(w)_{\max} - [\mathbf{D}(w, G)]_{ij} & \text{if } i \neq j \\ Vw(w)_i & \text{if } i = j \end{cases} \quad (37)$$

where $[D]_{ij}$ is the ij th element of the distance matrix D which is equal to the graph distance between vertices v_i and v_j .

4.5.19

The Reciprocal Reverse Wiener Matrix

The reciprocal reverse Wiener matrix $RRW(w, G)$ of a vertex- and edge-weighted molecular graph G with N vertices is the square $N \times N$ symmetric matrix defined

with the equation [35, 44]:

$$[\mathbf{RRW}(w, G)]_{ij} = \begin{cases} 1/[\mathbf{RW}(w, G)]_{ij} & \text{if } i \neq j \\ [\mathbf{RW}(w, G)]_{ii} & \text{if } i = j \end{cases} \quad (38)$$

4.5.20

The Szeged Matrix

Consider a graph G with N vertices labeled from 1 to N ; for any pair of vertices v_i and v_j from G , $v_i, v_j \in V(G)$, n_{ij} represents the number of vertices v_k of the molecular graph G having the property $td_{ki} < td_{kj}$ and n_{ji} represents the number of vertices v_k with the property $td_{kj} < td_{ki}$ [47]:

$$n_{ij} = |\{v_k : v_i, v_j, v_k \in V(G), td_{ki} < td_{kj}\}| \quad (39)$$

$$n_{ji} = |\{v_k : v_i, v_j, v_k \in V(G), td_{kj} < td_{ki}\}| \quad (40)$$

From the above two definitions it is clear that vertex v_i is counted in n_{ij} while vertex v_j is counted in n_{ji} ; if a vertex v_k is situated at the same topological distance from vertices v_i and v_j , i.e. $td_{ki} = td_{kj}$, the vertex is not counted neither in n_{ij} nor in n_{ji} . For any pair of vertices v_i and v_j from G , n_{ij} gives the number of vertices closer to vertex v_i and n_{ji} gives the number of vertices closer to vertex v_j . Diudea used the n_{ij} and n_{ji} numbers to define three new molecular matrixes [47]. For a graph G with N vertices, the Szeged matrix $Sz_u = Sz_u(G)$, is a square $N \times N$ non-symmetric matrix defined by [47]:

$$[Sz_u]_{ij} = \begin{cases} n_{ij} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \quad (41)$$

The non-symmetric Szeged matrix $Sz_u(44)$ of 1,2-dimethylcyclobutane 44 is:

	1	2	3	4	5	6
1	0	3	2	4	5	3
2	3	0	4	2	3	5
3	1	2	0	3	4	2
4	2	1	3	0	2	4
5	1	1	2	1	0	3
6	1	1	1	2	3	0

The path Szeged matrix $Sz_p = Sz_p(G)$ is a square $N \times N$ symmetric matrix obtained from Sz_u through a symmetrization operation [47]:

$$[Sz_p]_{ij} = \begin{cases} n_{ij}n_{ji} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases} \quad (42)$$

The symmetric path Szeged matrix $Sz_p(44)$ of 1,2-dimethylcyclobutane is:

$$Sz_p(44)$$

	1	2	3	4	5	6
1	0	9	2	8	5	3
2	9	0	8	2	3	5
3	2	8	0	9	8	2
4	8	2	9	0	2	8
5	5	3	8	2	0	9
6	3	5	2	8	9	0

The edge Szeged matrix $Sz_e = Sz_e(G)$ is a square $N \times N$ symmetric matrix obtained from Sz_u through a similar symmetrization operation [47]:

$$[Sz_e]_{ij} = \begin{cases} n_{ij}n_{ji} & \text{if } i \neq j \text{ and } e_{ij} \in E(G) \\ 0 & \text{if } i = j \text{ or } e_{ij} \notin E(G) \end{cases} \quad (43)$$

where $E(G)$ represents the edge set of G . An alternative definition of the Sz_e matrix is:

$$[Sz_e]_{ij} = [A]_{ij}n_{ij}n_{ji} = [A]_{ij}[Sz_p]_{ij} \quad (44)$$

where $[A]_{ij}$ is the element of the adjacency matrix $A = A(G)$. The edge Szeged matrix of 1,2-dimethylcyclobutane, $Sz_e(44)$, is:

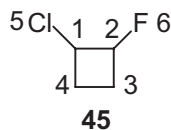
$$Sz_e(44)$$

	1	2	3	4	5	6
1	0	9	0	8	5	0
2	9	0	8	0	0	5
3	0	8	0	9	0	0
4	8	0	9	0	0	0
5	5	0	0	0	0	0
6	0	5	0	0	0	0

The edge Szeged matrix $Sz_e(w) = Sz_e(w, G)$ of a VEW molecular graph G with N vertices is the square $N \times N$ real symmetric matrix whose element $[Sz_e(w)]_{ij}$ is [35]:

$$[Sz_e(w, G)]_{ij} = \begin{cases} Vw(w)_i & \text{if } i = j \\ n_{ij}n_{ji}Ew(w)_{ij} & \text{if } e_{ij} \in E(G) \\ 0 & \text{if } e_{ij} \notin E(G) \end{cases} \quad (45)$$

where $Vw(w)_i$ is the weight of the vertex v_i , $Ew(w)_{ij}$ is the weight of the edge e_{ij} , and w is the weighting scheme used to compute the parameters Vw and Ew .



The edge Szeged matrix Sz_e with edge weights obtained from the atomic polarizability weighting scheme P , is given for 1-chloro-2-fluorocyclobutane 45:

$Sz_e(P, 45)$

	1	2	3	4	5	6
1	0	9.000	0	8.000	4.037	0
2	9.000	0	8.000	0	0	15.799
3	0	8.000	0	9.000	0	0
4	8.000	0	9.000	0	0.000	0
5	4.037	0	0	0	0.193	0
6	0	15.799	0	0	0	-2.160

4.6 Molecular Graph Polynomials

Polynomials and spectra derived from molecular graphs and matrixes have important applications in chemistry in connection with the molecular orbital theory of unsaturated compounds [4, 7, 17–19, 48]. The characteristic and acyclic polynomials of conjugated molecules were used to define several aromaticity indices. Various topological indices and other QSAR descriptors were derived from graph polynomials, spectra, and spectral moments. We present here the molecular graph polynomials that are used in the definition of structural descriptors and topological indices.

4.6.1 The Characteristic Polynomial

The characteristic polynomial $Ch(G, x)$ of a molecular graph G is the characteristic polynomial of its adjacency matrix $A(G)$ [4, 7, 48]:

$$Ch(G, x) = \det(xI - A) = \sum_{n=0}^N c_n x^{N-n} \quad (46)$$

where I is the unit matrix of order N and c_n is the n th coefficient of the characteristic polynomial. A graph eigenvalue x_i is a zero of the characteristic polynomial:

$$Ch(G, x_i) = 0 \quad (47)$$

for $I = 1$ to N . The complete set of graph eigenvalues x_1, x_2, \dots, x_N forms the spectrum of the graph G , $Sp(A, G) = (x_i, i = 1, 2, \dots, N)$. The characteristic polynomials and spectra for several linear and cyclic graphs are presented below:

$$Ch(L_0) = 1$$

$$Ch(L_1) = x$$

$$Ch(L_2) = x^2 - 1$$

$$Ch(L_3) = x^3 - 2x$$

$$Ch(L_4) = x^4 - 3x^2 + 1$$

$$Ch(L_5) = x^5 - 4x^3 + 3x$$

$$Ch(L_6) = x^6 - 5x^4 + 6x^2 - 1$$

$$Ch(R_3) = x^3 - 3x - 2$$

$$Ch(R_4) = x^4 - 4x^2$$

$$Ch(R_5) = x^5 - 5x^3 + 5x - 2$$

$$Ch(R_6) = x^6 - 6x^4 + 9x^2 - 4$$

$$Sp(A, L_2) = \{1, -1\}$$

$$Sp(A, L_3) = \{1.414\ 21, 0, -1.414\ 21\}$$

$$Sp(A, L_4) = \{1.618\ 03, 0.618\ 03, -0.618\ 03, -1.618\ 03\}$$

$$Sp(A, L_5) = \{1.732\ 05, 1, 0, -1, -1.732\ 05\}$$

$$Sp(A, L_6) = \{1.801\ 94, 1.246\ 98, 0.445\ 04, -0.445\ 04, -1.246\ 98, -1.801\ 94\}$$

$$Sp(A, R_3) = \{2, -1, -1\}$$

$$Sp(A, R_4) = \{2, 0, 0, -2\}$$

$$Sp(A, R_5) = \{2, 0.618\ 03, 0.618\ 03, -1.618\ 03, -1.618\ 03\}$$

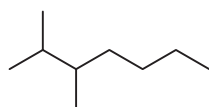
$$Sp(A, R_6) = \{2, 1, 1, -1, -1, -2\}$$

The spectral moment of order k , SM_k , is defined as:

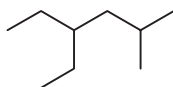
$$SM_k = \sum_{i=1}^N x_i^k = \text{Tr } A^k = \sum_{i=1}^N [A^k]_{ii} \quad (48)$$

An important result is the geometric interpretation of Eq. (48). The number of self-returning walks of length k may be computed by considering the diagonal elements of the first k powers of the adjacency matrix A , because each diagonal element $[A^k]_{ii}$ of the matrix A^k can be interpreted as the sum of all self-returning walks of length k from/to vertex v_i .

It was initially conjectured that the characteristic polynomial and its spectrum might be used as unique descriptors of graphs; however, non-isomorphic graphs with identical spectrum, spectral moments and characteristic polynomial were found, and they were called isospectral or cospectral graphs.



46



47

The pair of isospectral 4-trees representing 2,3-dimethylheptane **46** and 3-ethyl-5-methylhexane **47** has the same characteristic polynomial:

$$Ch(\mathbf{46}) = Ch(\mathbf{47}) = x^9 - 8x^7 + 19x^5 - 14x^3 + 2x$$

4.6.2

The Acyclic (Matching) Polynomial

The acyclic (matching) polynomial of a graph G is defined by the equation [48]:

$$Ac(G, x) = \sum_{k=0}^{N/2} (-1)^k m(G, k) x^{N-2k} = \sum_{n=0}^N a_n x^{N-n} \quad (49)$$

where $m(G, k)$ denotes the number of k -matchings of G . This polynomial plays an important role in statistical physics (theory of monomer–dimer systems), in developing topological indices for quantitative structure–property relationships, and in quantum organic chemistry (topological resonance energy).

4.6.3

The Characteristic Polynomial of a Molecular Matrix

In the previous section we introduced the most important molecular graph matrixes, such as the adjacency, distance, resistance distance, or Szeged matrixes. Any symmetric molecular matrix can be used to derive a corresponding characteristic polynomial and matrix spectrum, which are the basis of computing various topological indices for QSAR. The characteristic polynomial of the molecular matrix M of the graph G is:

$$Ch(M, G) = \det(xI - M) \quad (50)$$

The spectrum of the molecular matrix M , $Sp(M, G) = (x_i, i = 1, 2, \dots, N)$, is formed by the complete set of eigenvalues of M .

4.7

Enumeration of Kekulé Structures

The number of Kekulé structures (Kekulé structure count, KSC) of a conjugated hydrocarbon gives information on the thermodynamic stability and chemical reactivity and is used for the computation of various structural indices for benzenoid hydrocarbons [4, 7, 19–22]: the Pauling Bond Order, PBO ; the stability index SI , $SI = KSC^{2/N}$; the π -electron energy. Kekulé structures are considered in valence-bond resonance-theoretical models such as the Pauling–Wheland model, the Simpson–Herndon model, and the conjugated-circuits model. The calculation of the KSC can be performed with recurrence relationships, matrix methods, and

explicit combinatorial expressions derived for a large number of classes of conjugated hydrocarbons: various classes of cata-condensed benzenoid hydrocarbons, honeycomb lattice strips, polymers.

Some general expressions and recurrence relationships which allow the computation of KSC for benzenoid or general conjugated molecular graphs are presented. The determinant of the adjacency matrix A of the Hückel graph G of a benzenoid hydrocarbon is related to the number of Kekulé structures in G [2]:

$$\det \mathbf{A}(G) = (-1)^{N/N} KSC(G)^2 \quad (51)$$

If G is the Hückel graph of a conjugated hydrocarbon then [4, 7, 19]:

$$KSC(G) = KSC(G - e_{ij}) + KSC(G - v_i - v_j) \quad (52)$$

When the vertex v_i is of degree one, Eq. (52) becomes [4, 7, 19]:

$$KSC(G) = KSC(G - v_i - v_j) \quad (53)$$

If the conjugated system G is an essentially disconnected benzenoid composed of two non-interacting fragments G_1 and G_2 , then [4, 7, 19]:

$$KSC(G) = KSC(G_1)KSC(G_2) \quad (54)$$

If G is the Hückel graph of a benzenoid hydrocarbon with N vertices, then the free term of the characteristic polynomial, i.e. the coefficient of x^0 , $Ch(G, 0)$, is related to $KSC(G)$ [4, 7, 19]:

$$Ch(G, 0) = (-1)^{N/2} KSC(G)^2 \quad (55)$$

If G is the Hückel graph of a conjugated hydrocarbon with N vertices, then the free term of the acyclic polynomial, i.e. the coefficient of x^0 , $Ac(G, 0)$, is connected with $KSC(G)$ [4, 7, 19]:

$$Ac(G, 0) = (-1)^{N/2} KSC(G) \quad (56)$$

If G is the Hückel graph of a benzenoid hydrocarbon, then:

$$KSC(G - e_{ij})KSC(G - v_i - v_j) = \sum_{k=1}^r KSC(G - C_k)^2 \quad (57)$$

where C_i is a cycle of G and the summation goes over all r cycles in G containing the edge e_{ij} . If G is the Hückel graph of a benzenoid hydrocarbon then $KSC(G)$ can be expressed as a function of the number of Kekulé structures of the subgraphs of G which do not contain the edge e_{ij} [49]:

$$KSC(G)^2 = KSC(G - e_{ij})^2 + KSC(G - v_i - v_j)^2 + 2 \sum_{k=1}^r KSC(G - C_k)^2 \quad (58)$$

where the summation goes over all r cycles in G which contain the edge e_{ij} .

If G is the Hückel graph of a benzenoid hydrocarbon then $KSC(G)$ is related to the number of Kekulé structures of the subgraphs of G obtained after the removal of a vertex v_i [49]:

$$KSC(G)^2 = \sum_{j=1}^{deg_i} KSC(G - v_i - v_j)^2 + 2 \sum_{k=1}^r KSC(G - C_k)^2 \quad (59)$$

where the first summation goes over all deg_i vertices adjacent to vertex v_i , and the second summation goes over all r cycles C_k which contain vertex v_i . If G is the Hückel graph of a conjugated hydrocarbon then $KSC(G)$ can be computed from the number of Kekulé structures of the subgraphs of G corresponding to the decomposition at its vertex v_i [49]:

$$KSC(G) = \sum_{j=1}^{deg_i} KSC(G - v_i - v_j) \quad (60)$$

where the summation goes over all deg_i vertices adjacent to vertex v_i .

The Pauling Bond Order PBO_{ij} between two adjacent vertices v_i and v_j from the Hückel graph of a benzenoid hydrocarbon can be computed with the equation [19, 20]:

$$PBO_{ij} = KSC(G - v_i - v_j) / KSC(G) = 1 - KSC(G - e_{ij}) / KSC(G) \quad (61)$$

The Pauling bond order was successfully correlated with experimentally determined bond lengths of various benzenoid hydrocarbons.

4.8

Molecular Graphs and Hückel Molecular Orbital Theory

The Hückel theory, the first molecular orbital theory used in organic chemistry, provided qualitative information on the structure and reactivity of conjugated compounds. There is a close connection between Hückel theory and the graph spectral theory of Hückel molecular graphs: the energies (in β units) of Hückel molecular orbitals of a conjugated molecule are identical with the spectrum of the adjacency matrix of the Hückel molecular graph of the respective molecule. Various quantities obtained from the Hückel molecular orbitals and energies (total π -electron energy, charge density, bond order, free valence, absolute hardness, etc.) can be derived from the spectral analysis of the Hückel molecular graph or from empirical equations relating them to selected graph invariants [4, 7, 17, 18].

4.9

The Topological Resonance Energy

The definition of the resonance energy is:

$$RE = E_{\pi, \text{mol}} - E_{\pi, \text{ref}} \quad (62)$$

where $E_{\pi, \text{mol}}$ and $E_{\pi, \text{ref}}$ are the π -electron energies of a conjugated molecule and the corresponding reference structure, respectively. Both E_{π} terms can be computed at different levels of theory, from Hückel to ab initio methods. There are many proposals on the definition of the reference structure, which is a hypothetical structure containing the same number of π -electrons as the corresponding conjugated molecule but without conjugation. The topological resonance energy (TRE) defined by Aihara [50] and by Gutman, Milun, and Trinajstić [51] is computed at the Hückel level:

$$TRE = \sum_{i=1}^N g_i (x_i - y_i) \quad (63)$$

where x_i is the i th value from the characteristic polynomial spectrum, y_i is the i th value from the acyclic polynomial spectrum, and g_i is the electron occupancy of the i th energy level. The TRE can be normalized by dividing its value to the total number of π -electrons, N_e , giving the TRE per π -electron (TREPE) value:

$$TREPE = TRE/N_e \quad (64)$$

TREPE represents the conjugation stabilization or destabilization that one π -electron contributes to the molecular π system. TRE values were computed for a large variety of conjugated compounds, including conjugated ions, radicals, and ion radicals, coumarins, thiocoumarins, and bridged heteroannulenes [7]. Selected TRE and TREPE values for conjugated hydrocarbons are presented in Table 4-5.

Tab. 4-5 TRE and TREPE values for conjugated hydrocarbons (in β units)

<i>molecule</i>	<i>TRE</i>	<i>TREPE</i>
benzene	0.276	0.046
naphthalene	0.390	0.039
anthracene	0.476	0.034
phenanthrene	0.546	0.039
cyclobutadiene	-1.228	-0.307
benzocyclobutadiene	-0.392	-0.049

4.10 Isomer Enumeration

Isomers are stable chemical compounds with an identical molecular formula but different structure, conformation or configuration which display different physicochemical properties. The isomer enumeration represents an important graph theory application in chemistry; the presentation of the most important isomer enumeration algorithms can be found in a number of reviews [52, 53]. Cayley developed generating functions for enumeration of alkanes and alkyl radicals and produced extensive numerical data on the number of isomers with various molecular formulas. Pólya introduced the most powerful isomer enumeration algorithm by using the molecular symmetry, weighting factors, and generating functions [52]. The mathematical background of the Pólya enumeration method is presented here, together with an example.

An arrangement of a set of objects is an ordering of these objects. A permutation π is an operation that changes one arrangement into another arrangement and can be represented in a two-row notation by the expression:

$$\pi = \begin{pmatrix} 1 & 2 & 3 & \cdots & i & \cdots & N \\ p_1 & p_2 & p_3 & \cdots & p_i & \cdots & p_N \end{pmatrix} \quad (65)$$

with the meaning that object 1 is permuted to object p_1 , object 2 is permuted to object p_2 , object 3 is permuted to object p_3 , object i is permuted to object p_i , and object N is permuted to object p_N . A permutation of a set of n objects is called a cyclic permutation (or cycle) of length m if it moves the object in position p_1 to position p_2 , the object in position p_2 to position p_3, \dots , the object in position p_{m-1} to position p_m , the object in position p_m to position p_1 , and all other objects are left fixed. A cyclic permutation is denoted by $((p_1, p_2, p_3, \dots, p_{m-1}, p_m))$. A transposition is a cycle of length 2. Consider four distinct objects a_1, a_2, a_3 , and a_4 and four arrangements with the four objects: $A_1 = (a_1 a_2 a_3 a_4)$; $A_2 = (a_1 a_3 a_2 a_4)$; $A_3 = (a_2 a_3 a_1 a_4)$; $A_4 = (a_2 a_3 a_4 a_1)$. The permutation which changes A_1 to A_2 is the transposition $\pi_1 = ((2\ 3))$; the permutation which changes A_1 to A_3 is a cycle of length 3 denoted $\pi_2 = ((1\ 2\ 3))$; the permutation which changes A_1 to A_4 is a cycle of length 4 denoted $\pi_3 = ((1\ 2\ 3\ 4))$. If we indicate also the objects that remain fixed in a permutation, the first two permutations are denoted: $\pi_1 = ((1)(2\ 3)(4))$; $\pi_2 = ((1\ 2\ 3)(4))$. The set of all possible permutations that can be applied to a set of n objects, together with the composition operation, forms a permutation group. If one considers a set containing three objects then the six permutations of its group, denoted S_3 , are: $p_1 = ((1)(2)(3))$, $p_2 = ((1\ 2)(3))$, $p_3 = ((1\ 3)(2))$, $p_4 = ((1)(2\ 3))$, $p_5 = ((1\ 2\ 3))$, $p_6 = ((1\ 3\ 2))$. For a group of n objects the set of all possible permutations corresponds to the group S_n , the symmetric group of degree n , containing $n!$ permutations.

Two permutations are disjoint if they act on mutually exclusive sets of objects in an arrangement. To any permutation of n objects π one assigns a monomial $s(\pi)$ in

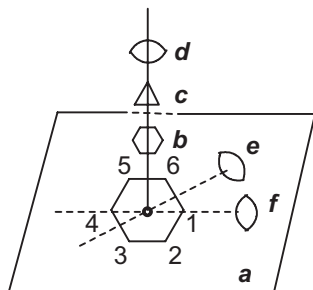


Fig. 4-1 The symmetry axes of benzene

the variable s_k corresponding to a cyclic permutation of length k in the unique product of disjoint cycles of π . A fixed object corresponds to a factor s_1^1 , m fixed objects correspond to s_1^m , and a transposition corresponds to s_2^1 . The factors associated with the above permutations π_1, π_2 , and π_3 are $s_1^2 s_2^1, s_1^1 s_3^1$, and s_4^1 .

Let Γ be a permutation group of degree n . For each permutation $\pi \in \Gamma$, let $j_k(\pi)$ be the number of cycles of length k in the disjoint cycle decomposition of π . The cycle index $Z(\Gamma)$ is the polynomial in n variable s_1, s_2, \dots, s_n given by the formula:

$$Z(\Gamma) = \frac{1}{|\Gamma|} \sum_{\pi \in \Gamma} \prod_{k=1}^n s_k^{j_k(\pi)} \quad (66)$$

4.10.1

Pólya's Theorem

If $c(x)$ is the counting series for a collection of weighted objects, and Γ is a permutation group acting on n positions such that Γ defines an equivalence relation on arrangements consisting of n objects, then the substituted cycle index $Z(\Gamma, c(x))$ is the counting series for nonequivalent arrangements consisting of n objects where the weight of an arrangement is the sum of the weights of the n objects of which it is composed.

The benzene cycle index derivation considers the symmetry elements indicated in Figure 4-1, corresponding to the permutations and cycle index terms presented in Table 4-6.

The cycle index for benzene is:

$$Z = \frac{1}{12} s_1^6 + 2s_6^1 + 2s_3^2 + 4s_2^3 + 3s_1^2 s_2^2 \quad (67)$$

The number of benzene substitution isomers with formula $C_6H_{6-k}X_k$ is obtained by substituting Eq. (87) in the cycle index for benzene, Eq. (67). This substitution gives the counting polynomial:

Tab. 4-6 Generation of the cycle index of benzene

Symmetry operation	1 2 3 4 5 6	Permutation	Cycle index term
E (a)	1 2 3 4 5 6	(1)(2)(3)(4)(5)(6)	s_1^6
C_6^+ (b)	2 3 4 5 6 1	(123456)	s_6^1
C_6^- (b)	6 1 2 3 4 5	(165432)	s_6^1
C_3^+ (c)	3 4 5 6 1 2	(135)(246)	s_3^2
C_3^- (c)	5 6 1 2 3 4	(153)(264)	s_3^2
C_2 (d)	4 5 6 1 2 3	(14)(25)(36)	s_2^3
C_2 (e)	2 1 6 5 4 3	(12)(36)(45)	s_2^3
C_2 (e)	6 5 4 3 2 1	(16)(25)(34)	s_2^3
C_2 (e)	4 3 2 1 6 5	(14)(23)(56)	s_2^3
C_2 (f)	1 6 5 4 3 2	(1)(4)(26)(35)	$s_1^2 s_2^2$
C_2 (f)	3 2 1 6 5 4	(2)(5)(13)(46)	$s_1^2 s_2^2$
C_2 (f)	5 4 3 2 1 6	(3)(6)(15)(24)	$s_1^2 s_2^2$

$$Z = (1/12)[(1+x)^6 + 2(1+x^6) + 2(1+x^3)^2 + 4(1+x^2)^3 + 3(1+x)^2(1+x^2)^2] = 1 + x + 3x^2 + 3x^3 + 3x^4 + x^5 + x^6 \quad (68)$$

showing that benzene has one unsubstituted isomer, one monosubstituted isomer, three disubstituted isomers, three trisubstituted isomers, etc.

4.11 Conclusions

The graph representation of chemical objects, their relationships, and their transformations has a profound impact in our understanding of the relationships between chemical structure and the physical, chemical, and biological properties of chemicals. Computational methods based on molecular graphs are used to encode chemicals and reactions in databases, to enumerate all isomers with a given molecular formula, or to generate combinatorial libraries. The adjacency matrix is familiar to chemists because they were first applied in Hückel molecular orbital theoretical calculations. In addition to the adjacency matrix, various other types of matrix can be associated with molecular graphs; these molecular matrixes represent important sources of topological indices. The rich literature on topological indices and their applications in QSPR, QSAR, and virtual screening of combinatorial libraries, shows that the most active application of graphs in chemistry is the development of new structural descriptors derived from molecular graphs. Significant developments were made by defining parameters for heteroatoms and multiple bonds, by introducing new molecular matrixes, and by computing topological indices with graph operators. For the future, major progress is expected in this

direction, with benefits for the drug design process and for structure–property models.

References

- 1 F. HARARY, *Graph Theory*, Addison–Wesley, Reading, MA, 1971.
- 2 D. M. CVETKOVIĆ, M. DOOB, H. SACHS, *Spectra of Graphs. Theory and Applications, 3rd edition*, Johann Ambrosius Barth, Heidelberg, 1995.
- 3 A. T. BALABAN (Ed.), *Chemical Applications of Graph Theory*; Academic Press, London, 1976.
- 4 I. GUTMAN, O. E. POLANSKY, *Mathematical Concepts in Organic Chemistry*, Springer, Berlin, 1986.
- 5 D. H. ROUVRAY (Ed.), *Computational Chemical Graph Theory*, Nova Science, New York, 1990.
- 6 D. BONCHEV, D. H. ROUVRAY (Eds.), *Chemical Graph Theory. Introduction and Fundamentals*, Abacus Press/Gordon and Breach Science Publishers, New York, 1991.
- 7 N. TRINAJSTIĆ, *Chemical Graph Theory, 2nd ed.*, CRC Press, Boca Raton, 1992.
- 8 A. T. BALABAN (Ed.), *From Chemical Topology to Three-Dimensional Geometry*, Plenum, New York, 1997.
- 9 L. B. KIER, L. H. HALL, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976.
- 10 A. T. BALABAN, A. CHIRIAC, I. MOTOC, Z. SIMON, *Steric Fit in Quantitative Structure–Activity Relations*, Lect. Notes Chem. Vol. 15, Springer, Berlin, 1980.
- 11 D. BONCHEV, *Information Theoretic Indices for Characterization of Chemical Structure*, Research Studies Press – Wiley, Chichester, UK, 1983.
- 12 L. B. KIER, L. H. HALL, *Molecular Connectivity in Structure–Activity Analysis*, Research Studies Press, Letchworth, 1986.
- 13 N. VOICULETZ, A. T. BALABAN, I. NICULESCU-DUVAZ, Z. SIMON, *Modeling of Cancer Genesis and Prevention*, CRC Press, Boca Raton, 1990.
- 14 L. B. KIER, L. H. HALL, *Molecular Structure Description: The Electrotological State*, Academic Press, San Diego, 1999.
- 15 J. DEVILLERS, A. T. BALABAN (Eds.), *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon and Breach Science Publishers, The Netherlands, 1999.
- 16 M. V. DIUDEA (Ed.), *QSPR/QSAR Studies by Molecular Descriptors*, Nova Science, Huntington, N.Y., 2001.
- 17 A. GRAOVAC, I. GUTMAN, N. TRINAJSTIĆ, *Topological Approach to the Chemistry of Conjugated Molecules*, Springer, Berlin, 1977.
- 18 J. R. DIAS, *Molecular Orbital Calculations Using Chemical Graph Theory*, Springer, Berlin, 1993.
- 19 S. J. CYVIN, I. GUTMAN, *Kekulé Structures in Benzenoid Hydrocarbons*, Lect. Notes Chem. Vol. 46, Springer, Berlin, 1988.
- 20 I. GUTMAN, S. J. CYVIN, *Introduction to the Theory of Benzenoid Hydrocarbons*, Springer, Berlin, 1989.
- 21 I. GUTMAN, S. J. CYVIN, *Advances in the Theory of Benzenoid Hydrocarbons*, Topics Curr. Chem. Vol. 153, Springer, Berlin, 1990.
- 22 S. J. CYVIN, J. BRUNVOLL, B. N. CYVIN, *Theory of Coronoid Hydrocarbons*, Lect. Notes Chem. Vol. 54, Springer, Berlin, 1991.
- 23 G. PÓLYA, R. C. READ, *Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds*, Springer, Berlin, 1987.
- 24 S. FUJITA, *Symmetry and Combinatorial Enumeration in Chemistry*, Springer, Berlin, 1991.
- 25 O. IVANCIUC, *Canonical Numbering and Constitutional Symmetry*, in: *The Encyclopedia of Computational Chemistry*, P. v. R. SCHLEYER, N. L.

- ALLINGER, T. CLARK, J. GASTEIGER, P. A. KOLLMAN, H. F. SCHAEFER III, AND P. R. SCHREINER (Eds.), John Wiley and Sons, Chichester, 1998, pp. 167–183.
- 26 O. N. TEMKIN, A. V. ZEIGARNIK, D. BONCHEV, *Chemical Reaction Networks. A Graph-Theoretical Approach*, CRC Press, Boca Raton, 1996.
- 27 J. KOČA, M. KRATOCHVÍL, V. KVASNIČKA, L. MATYSKA, J. POSPÍCHAL, *Synthon Model of Organic Chemistry and Synthesis Design, Lect. Notes Chem. Vol. 51*, Springer, Berlin, 1989.
- 28 M. BARYSZ, G. JASHARI, R. S. LALL, V. K. SRIVASTAVA, AND N. TRINAJSTIĆ, *On the Distance Matrix of Molecules Containing Heteroatoms*. In: *Chemical Applications of Topology and Graph Theory*, R. B. KING (Ed.), Elsevier, Amsterdam, 1983, pp. 222–227.
- 29 O. IVANCIUC, *Rev. Roum. Chim.* **2000**, 45, 289–301.
- 30 F. R. BURDEN, *J. Chem. Inf. Comput. Sci.* **1989**, 29, 225–227.
- 31 R. S. PEARLMAN AND K. M. SMITH, *J. Chem. Inf. Comput. Sci.* **1999**, 39, 28–35.
- 32 O. IVANCIUC, *Rev. Roum. Chim.* **2001**, 46, 1047–1066.
- 33 N. TRINAJSTIĆ, D. BABIĆ, S. NIKOLIĆ, D. PLAVŠIĆ, D. AMIĆ, AND Z. MIHALIĆ, *J. Chem. Inf. Comput. Sci.* **1994**, 34, 368–376.
- 34 O. IVANCIUC, *Rev. Roum. Chim.* **2001**, 46, 1331–1347.
- 35 O. IVANCIUC, *J. Chem. Inf. Comput. Sci.* **2000**, 40, 1412–1422.
- 36 O. IVANCIUC, T.-S. BALABAN, AND A. T. BALABAN, *J. Math. Chem.* **1993**, 12, 309–318.
- 37 O. IVANCIUC AND A. T. BALABAN, *MATCH (Commun. Math. Chem.)* **1994**, 30, 141–152.
- 38 M. RANDIĆ, *J. Chem. Inf. Comput. Sci.* **1997**, 37, 1063–1071.
- 39 O. IVANCIUC, *Rev. Roum. Chim.* **2000**, 45, 587–596.
- 40 D. J. KLEIN AND M. RANDIĆ, *J. Math. Chem.* **1993**, 12, 81–95.
- 41 O. IVANCIUC, *ACH – Models Chem.* **2000**, 137, 607–631.
- 42 M. V. DIUDEA, *J. Chem. Inf. Comput. Sci.* **1996**, 36, 535–540.
- 43 M. RANDIĆ, *New J. Chem.* **1997**, 21, 945–951.
- 44 O. IVANCIUC, T. IVANCIUC, A. T. BALABAN, *Internet Electron. J. Mol. Des.* **2002**, 1, 467–487. <http://www.biochempress.com>.
- 45 O. IVANCIUC, T. IVANCIUC, A. T. BALABAN, *ACH – Models Chem.* **2000**, 137, 57–82.
- 46 A. T. BALABAN, D. MILLS, O. IVANCIUC, AND S. C. BASAK, *Croat. Chem. Acta* **2000**, 73, 923–941.
- 47 M. V. DIUDEA, *J. Chem. Inf. Comput. Sci.* **1997**, 37, 292–299.
- 48 O. IVANCIUC, T. IVANCIUC, M. V. DIUDEA, *Roum. Chem. Quart. Rev.* **1999**, 7, 41–67.
- 49 O. IVANCIUC, A. T. BALABAN, *J. Math. Chem.* **1992**, 11, 169–177.
- 50 J. AIHARA, *J. Am. Chem. Soc.* **1976**, 98, 2750–2758.
- 51 I. GUTMAN, M. MILUN, N. TRINAJSTIĆ, *J. Am. Chem. Soc.* **1977**, 99, 1692–1704.
- 52 G. PÓLYA, R. C. READ, *Combinatorial Enumeration of Groups, Graphs, and Chemical Compounds*, Springer, Berlin, 1987.
- 53 A. T. BALABAN, J. W. KENNEDY, L. V. QUINTAS, *J. Chem. Educ.* **1988**, 65, 304–313.
- 54 O. IVANCIUC, A. T. BALABAN, *Graph Theory in Chemistry*, in: *The Encyclopedia of Computational Chemistry*, P. v. R. SCHLEYER, N. L. ALLINGER, T. CLARK, J. GASTEIGER, P. A. KOLLMAN, H. F. SCHAEFER III, AND P. R. SCHREINER (Eds.), John Wiley & Sons, Chichester, 1998, pp. 1169–1190.