GEOMETRICAL DESCRIPTORS OF MOLECULES

Pavel TVARŮŽEK and Jaro KOMENDA

Department of Theoretical and Physical Chemistry, Masaryk University, 611 37 Brno

Received March 9, 1990
Accepted July 9, 1990

Individual molecules and/or their conformers are unequivocally described by the geometrical descriptors derived from the proper values of the square matrix belonging to the *K* graph. This graph is constructed on the basis of the bonds existing in the molecule, their relative lengths, and dihedral angles (if the latter assume the values of 0, 90, 180, and 270°).

The QSPR (Quantitative Structure Property Relationships) studies start from descriptors assigned to individual compounds. As a rule, the descriptors are constructed on the basis of the respective structural formula which gives information about qualitative and quantitative composition of the molecule as well as about its structure. For instance, the linear nomenclature[1] which satisfactorily describes the chemical structure is well accessible for computer treatment.

However, attempts at representing a chemical structure by means of only numerical characteristics[2] appeared as early as in the 19th century. One possible way of solving such problems starts from a transformation of a chemical formula into the respective chemical multigraph which can be represented by means of the corresponding adjacency matrix and distance matrix. The former one is a square matrix of the n-th order where n means the number of apexes in the multigraph, its $a_{i,i}$ elements being (by definition) equal to zero, and an element $a_{i,j} = 1$ only if there exists an edge ncident with the i and j apexes, and $a_{i,j} = 0$ in the other cases. The distance matrix iis a square matrix of the n-th order where n means the number of apexes in the multigraph, and its $d_{i,j}$ element is equal to the shortest pathway length from the apex i to the apex j. The minimum path length between the apexes is defined as the minimum sum of the distances between individual apexes on the pathway from i to j. The distance between the apexes connected by an edge and the magnitude of the $d_{i,i}$ elements is evaluated nonuniformly but always with regard to the individuality of the atoms which exist in the compound described[2,4,6]. These matrices then make it possible to calculate a series of topological indices and/or their informational analogues[3,4]. As the chemical multigraphs are predominantly planar, they are irrelevant to the three-dimensional structure of chemical compounds. The spatial structure is satisfactorily represented e.g. by the parameters used in the QSAR

(Quantitative Structure Activity Relationships) analysis by Hansch[5], however, the parametrization of QSAR is known for a limited number of molecules only. Another possibility of determination of geometrical descriptors for any molecular structure is provided by molecular mechanics. From their results there follows the calculation of geometrical descriptors as the matrix characteristics of the inertia tensor[6]. Less complicated geometrical descriptors are obtained from the values of the steric constants by Taft or possibly from the van der Waals radii of the atoms forming this molecular structure[7].

All the given descriptors assume the same values for different conformers, optical, as well as geometrical isomers. Therefore, we will define descriptors relevant to the spatial (stereoisomeric) structure of molecule.

Definition and Properties of Geometrical Descriptors

A structural formula of molecule is assigned a graph K obeying two rules: 1. Each bond in the structural formula is represented by an apex in the graph K. 2. In the graph K the edges are defined whose terminal apexes correspond to the bonds for which the dihedral angle assumes the values of 0, 90, 180, 270, or 360°.

The graph K is not oriented, disconnected, and contains circles in some cases. Each apex of the graph K is evaluated by a dimensionless number equal to the ratio of the length of corresponding covalent bond to the C—C bond length (0·154 nm). Hence, the apex corresponding to the C—H bond is evaluated as 0·7, and the evaluation of the apexes corresponding to the bonds between carbon atoms of different hybridization is as follows:

$C(sp^3)$—$C(sp^3)$	1·0	$C(sp^2)$—$C(sp^2)$	0·9
$C(sp^3)$—$C(sp^2)$	1·0	$C(sp^2)$—$C(sp)$	0.8
$C(sp^3)$—$C(sp)$	0·9	$C(sp)$—$C(sp)$	0·8

The evaluation of edges of graph K is given by the spatial arrangement of the bonds in the structural formula of molecule. If the dihedral angle of the bonds corresponding to the terminal apexes of an edge is equal to 0° or 360°, then the edge is evaluated as 2π. In the case of the dihedral angle equal to 180° (the chemical bonds are opposite to each other) the bond evaluation is equal to π. For the dihedral angles of 90 and 270° the edge evaluations are $\pi/2$ and $3\pi/2$, respectively.

For a graph K a square matrix is defined as $[\boldsymbol{K}] = (k_{i,j})$, $i = 1, \ldots, n$ is the number of apexes in the graph K (and, hence, the number of chemical bonds in the molecule). The element $k_{i,i}$ is equal to the evaluation of the i-th apex of the graph K. The element $k_{i,j} = 0$ just only if there exists no edge incident with the apexes i and j. In all other cases the magnitude of the element $k_{i,j}$ is equal to the evaluation of the respective edge. The matrix $[\boldsymbol{K}]$ is symmetrical, its elements are non-negative

numbers. Therefore the set of the eigenvalues of matrix $[\boldsymbol{K}]$ $t_1, \ldots, t_n$ is the set of real numbers.

The characteristical polynomial of the matrix $[\boldsymbol{K}]$ is

$$P(t) = (-1)^n \det([\boldsymbol{K}] - [t][\boldsymbol{E}]), \tag{1}$$

where $[\boldsymbol{E}]$ is the unit matrix, or

$$P(t) = t^n + a_1 t^{n-1} + \ldots + a_n . \tag{2}$$

The relation (3) defines the descriptor Z of the molecule on the basis of values of the coefficient of this characteristical polynomial

$$Z = \sum_i |a_i| . \tag{3}$$

This is an analogy of the polynomial index defined for the adjacency matrix and the distance matrix of the graph[8].

As shown by Eq. (1) the graph K – and, hence, also the spatial structure of molecule – is related to the numerical characteristics based on the magnitude of the eigenvalues of the $[\boldsymbol{K}]$ matrix. Hence, it is also possible to define the descriptor S as the product

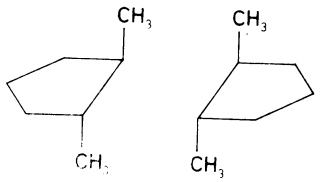
$$S = \prod_i t_i . \tag{4}$$

At the same time it is $S = a_n$ (ref.[9]).

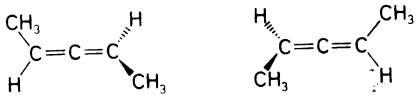
From properties of eigenvalues of a matrix it follows that the descriptor Z is independent of the numbering of apexes of graph K. Hence, it is its invariant and, therefore, is also invariant to the molecular structure characterized. However, the same numerical magnitude of Z can belong to two different graphs, i.e. the descriptor is degenerated. From Eq. (1) it follows that the descriptor Z is degenerated for the graphs $K1$ and $K2$ if the matrices $[\boldsymbol{K}_1]$ and $[\boldsymbol{K}_2]$ corresponding to these graphs have the same eigenvalues. Such a case is encountered either if the matrices $[\boldsymbol{K}_1]$ and $[\boldsymbol{K}_2]$ are identical or when these matrices have special properties. The matrices $[\boldsymbol{K}_1]$ and $[\boldsymbol{K}_2]$ are identical if the mutually corresponding elements of these matrices are equivalent. This situation is encountered also in the cases when the matrices contain insufficient amount of information about the chemical structures. In such cases the matrices are either diagonal or very sparse. It is also possible that the matrices are identical because the molecular structures cannot be differentiated by the described way of constructing the matrices. The matrix $[\boldsymbol{K}_1]$ has the same eigenvalues as the matrix $[\boldsymbol{K}_2]$ also in such case when the both matrices are similar. In general, however, nothing concrete can be said about the equality of eigenvalues of different matrices[9].

Applications

Table I presents the S and Z values calculated for some molecules. In the cases of rotation isomers of ethane and butane the values of S and Z descriptors differentiate well between the individual conformational structures. Both the descriptors calculated also discriminate between the boat and chair forms of cyclohexane and between



I *II*



III *IV*

TABLE I

Values of S and Z descriptors of some hydrocarbon molecules

Compound	Rotamer	S	Z
Ethane	*syn*-periplanar	−825·185	2109·24
	syn-clinal	−59265·9	127413
	with diagonal **K**	0·117649	43·0751
Butane	*anti*-periplanar	817089	$3{\cdot}947 \cdot 10^6$
	anti-clinal	−293954	$3{\cdot}389 \cdot 10^6$
	syn-clinal	$3{\cdot}203 \cdot 10^6$	$6{\cdot}303 \cdot 10^7$
	syn-periplanar	$2{\cdot}411 \cdot 10^6$	$2{\cdot}945 \cdot 10^7$
	with diagonal **K**	0·0282486	1602·8
Cyclohexane	boat	$2{\cdot}442 \cdot 10^{11}$	$1{\cdot}399 \cdot 10^{12}$
	chair	$-7{\cdot}048 \cdot 10^7$	$5{\cdot}187 \cdot 1g^{12}$
Methylcyclohexane	axial	$-1{\cdot}337 \cdot 10^{12}$	$8{\cdot}129 \cdot 10^{12}$
	equatorial	$5{\cdot}801 \cdot 10^{10}$	$1{\cdot}543 \cdot 10^{12}$
1,2-Dimethylcyclopentane	*I*	$-7{\cdot}684 \cdot 10^{13}$	$1{\cdot}485 \cdot 10^{14}$
	II	$-7{\cdot}548 \cdot 10^{13}$	$2{\cdot}072 \cdot 10^{14}$
2,3-Pentadiene	*III*	2083·5	343762
	IV	44898·5	$3{\cdot}886 \cdot 10^6$

the equatorial and axial conformers of methylcyclohexane. On the basis of magnitude of the *S* and *Z* descriptors it is possible to discriminate among rotamers for which the [*K*] matrices are diagonal. Different values of the *S* and *Z* descriptors are also observed with the chiral *trans* isomers of 1,2-dimethylcyclopentane (*I* and *II*) as well as with the optical antipodes of 2,3-pentadiene (*III* and *IV*). The [*K*] matrices constructed for the chiral molecules of 2-chloropropanoic acid contain such a small number of nonzero nondiagonal elements that they are identical for the two optically active compounds. Thus both the descriptors described are degenerated in this case.

Conclusion

The geometrical descriptors newly defined respect the hybridization states of atoms in the molecules described. Whereas the construction of [*K*] matrix is relatively exacting with regard to information, the numerical calculation of coefficients of the characteristical polynomial presents no algorithmic problems. The possibility of degeneration of the two descriptors depends immediately on the information content of the [*K*] matrix and on the properties of this matrix. It turns out that the descriptors defined can discriminate among conformers, optical isomers of chiral molecules with assymetrically arranged substituents on the atomic grouping, *cis* and *trans* geometrical isomers, diastereoisomers, and among compounds containing several adjacent asymmetrical carbon atoms (*threo* and *erythro* configurations).

REFERENCES

1. Smith E. G.: *The Wiswesser Line-Formula Chemical Notation.* McGraw-Hill, New York 1968.
2. King R. B. (Ed.): *Chemical Applications of Topology and Graph Theory.* Elsevier, New York 1983.
3. Slanina Z.: Chem. Listy *72*, 1 (1978).
4. Hladká E., Matyska L.: Chem. Listy *82*, 1009 (1988).
5. Franke R.: *Theoretical Drug Design Methods.* Akademie-Verlag, Berlin 1984.
6. Stuper A. J., Brugger W. E., Jurs P. C.: *Computer Assisted Studies of Chemical Structure and Biological Function.* Wiley, New York 1979.
7. Tichý M.: *Účinnost xenobiotik a chemická struktura.* Avicenum, Prague 1985.
8. Balaban A. T., Motoc I., Bonchev D., Mekenyan O. in: *Steric Effects in Drug Design* (F. L. Boschke, Ed.), p. 38. Akademie-Verlag, Berlin 1983.
9. Marcus M., Minc H.: *Obzor po teorii matric i matričnych neravenstv.* Nauka, Moscow 1972.

Translated by J. Panchartek.