Journal of Data Science **1**(2003), 361-389

The Matrix Expression, Topological Index and Atomic Attribute of Molecular Topological Structure

Qian-Nan Hu[1], Yi-Zeng Liang[1], and Kai-Tai Fang[2]
[1]*Central South University and* [2]*Hong Kong Baptist University*

Abstract: The matrix expression, topological index and atomic attribute of molecular topological structure are reviewed. Nine matrices, twenty-six kinds of indices and eight methods dealing with weighted molecular graphs are summed up in three tables. Some shortcomings of the topological indices are discussed as: (1) the physical-chemical meaning of topological index is not explicit; (2) it is difficult to interpret the QSAR and QSPR models derived from the topological indices; and (3) topological index usually neglects the stereochemical information or the three-dimensional structure of the molecule. Three directions of topological index are focused on: (1) description of local information; (2) studies on inter-correlation of topological index; and (3) variable index.

Key words: Atomic attribute, matrix expression, molecular topological structure, QSAR/QSPR, topological index.

1. Introduction

Quantitative structure-activity relationships (QSAR) and quantitative structure- activity relationships (QSPR) represent attempts to correlate activities or properties with structural descriptors of compounds. To correlate and predict physical, chemical and biological activity/property from molecular structure is a very important and an unsolved problem in theoretical and computational chemistry, environmental chemistry, medical chemistry, and life science as well. Indeed, many QSAR/QSPR articles (Suresh and Hansch 2001, Koh *et al.* 1998, David, A. C. 2000, David, F. V. 2000, Yu *et* *al.* 1999, Yuan and Parrill 2000, Geiss and Frazier 2001, Zhao *et al.* 1998, Shapiro *et al.* 1998, Christian *et al.* 1999, Liu *et al.* 2001, and Huang *et al.* 2002), reviews (Simona *et al.* 2000, Hansch *et al.* 2001, Katritzky 1996, Gao *et al.* 1999, Karmarkar and Khadikar 2000, and Kurup *et al.* 2001), and monographs (Karcher and Deviller 1990, Fujita and Timmerman 1995, Zhou and Wang 2001, Wang 1993, 1997, Diudea 2000, Kubinyi *et al.* 1998, and Rambon *et al.* 2000) were published on different fields.

The first step, also the most important step in QSAR/QSPR, is to numerically code the chemical structures of various molecules so as to build a correlation model between the chemical structures of various chemical compounds and the corresponding chemical and biological activities/properties. Thus, how to exactly transfer the chemical formula (or molecular graph) into numerical format has been a major task in QSAR/QSPR researches. There are many methods to quantify the molecular structures, in which topological index is the most popular since it can be obtained directly from molecular structures and rapidly computed for large numbers of molecules. A research (Basak *et al.* 1999) concluded that the topological index is the first effective choice in QSAR research. Recently, many articles (Cao and Yuan 2001, Balaban 1998, Bonchev 2000, Miguel *et al.* 2001, Agrawal 2001, Madan 1997, 1999, Bonchev 2001, Li *et al.* 2000, Rücker and Rücker 1999, and Erovnik 1999), reviews (Bono *et al.* 2001, Pogliani 2000, Randić and Zupan 2001, Estrada and Molina 2001, Diudea *et al.* 1995, Katritzky and Gordeeva 1993, and Schultz 2000), and monographs (Devillers and Balaban 1999, Kier and Hall 1999, Xin 1991, Xu and Hu 2001, Trinajstic 1992, King 1992, Gutman *et al.* 1991, and King and Rouvray 1987) gave systematic and comprehensive studies. Graph theory is nowadays a standard method of theoretical and computational chemistry (Harary 1969, and Cvetkovic *et al.* 1995) and a large number of references are available on its application in chemistry (Devillers and Balaban 1999, Kier and Hall 1999, Xin 1991, Xu and Hu 2001, Trinajstic 1992, King 1992, Gutman *et al.* 1991, King and Rouvray 1987, Balaban 1995, and Diudea and Ivanciuc 1995).

Numerous activities and properties of organic molecules depend on the presence of specific atoms and/or functional groups in their structures. The main aim of QSAR/QSPR is to link the structure of a molecule to a biological activity or a property by means of a statistics tools, which can be expressed mathematically as follows (Devillers 1999):

$$\text{A/P} = f(\text{molecular structure}) = f(\text{molecular descriptors}).$$

where A/P denotes the activity or property, which is essentially a chemical or a biological measurement value. The activity or property of a molecule can be commonly used normal boiling point, heat of formation, critical temperature, density, flash points, refractive index, chromatographic retention time, and octanol-water partition coefficient as well. Here, $f(\cdot)$ denotes a function, which depends on the molecular structure or molecular descriptors. In general, the model function, say $f(\cdot)$, can be linear or non-linear depending on different complexity of the data.

In order to evaluate structural similarity and diversity of the molecules and/or to build QSAR model as shown in the above equation, one need first to obtain the suitable numerical molecular descriptors associated with the molecular structure in QSAR/QSPR researches. There are many numerical molecular descriptors available in chemistry, including physical-chemical parameters, topological index, 3D descriptors and quantum chemical indices. However, in most cases, many chemists prefer to use topological index as molecular descriptors to evaluate toxicity, and predict biological activity (David, A. C. 2000, Liu *et al.* 2001, Basak *et al.* 1994, and Basak and Grunwald 1994), since the topological indices offer a simple way of measuring molecular branching, shape, size, cyclicity, symmetry, centricity and complexity.

The aim of this paper is to introduce the topological indices, which are essentially numerical molecular descriptors associated with the molecular structure. In order to make it easier for readers to understand the methods coding the chemical structures from the molecular graphs, three major forms are first classified as matrix expression, topological index and atomic attribute of molecular topological structure as well. Then, nine matrices expressing molecular structure, twenty-six kinds of topological indices and seven methods dealing with weighted molecular graphs are summed up in three tables for readers' convenient usage.

2. Three Methods Describing Graph Structure

Table 1: Graph matrices and their definitions

Name of matrix (reference)	Definition
The Adjacency Matrix (Lukovits 2000) A=A(G)	$[\mathrm{A}]_{ij} = 1$ if $i \neq j$ and $e_{ij} \in E(G)$ $= 0$ if $i = j$ or $e_{ij} \notin E(G)$ where e_{ij} is the edge formed by atoms i and j, $E(G)$ means the set of edges in the molecular graph.
Distance matrix (Rouvary 1986, Mihalic *et al.* 1992, and Bonchev and Tinajstic 1977) D=D(G)	$[\mathrm{D}]_{ij} = \min(l(p_{ij}))$ if $i \neq j$ $= 0$ if $i = j$, where $\min(l(p_{ij}))$ is the shortest path between atoms i and j.
Reciprocal matrix (Ivanciuc 1989, Plavsic *et al.* 1993, and Ivanciuc *et al.* 1993 RD=RD(G)	$[\mathrm{RD}]_{ij} = 1/\mathrm{D}_{ij}$ if $i \neq j$ $= 0$ if $i = j$, in which, the elements of RD is the reciprocal (excluding the zero elements) of the elements of D matrix.
Detour matrix (Ivanciuc and Balaban 1994, Amic and Trinajstic 1995, Diudea *et al.* 1998, Rücker 1998, and Mihalic 1997) Δ=DD(G)	$[\Delta]_{ij} = \max(l(p_{ij}))$ if $i \neq j$ $= 0$ if $i = j$ where $\max(l(p_{ij}))$ denotes the longest path between atoms i and j.
Edge-adjacency matrix (Estrada 1995, 1996, 1999, Estrada and Ramirez 1996, and Estrada *et al.* 1998) E=E(G)	$[\mathrm{E}]_{ij} = 1$ if there is a common node $= 0$ otherwise.
The Laplacian matrix (Mohar 1989, Mohar *et al.* 1993, Gutman *et al.* 1994, Trinajstic *et al.* 1994, and Ivanciuc 1993) L=DEG(G)-A(G)	$[\mathrm{L}]_{ij} = \deg_i$ if $i \neq j$ $= -1$ if $e_{ij} \in \mathrm{E(G)}$ $= 0$ if $e_{ij} \notin \mathrm{E(G)}$ where $\deg_i$ means the vertex degree of atom i.

Table 1 (contiuned): Graph matrices and their definitions

Name of matrix (reference)	Definition
The x matrix (Randić 1992) χ=X(G)	$\chi_{ij} = (\deg_i \deg_j)^{-1/2}$ if $e_{ij} \in$ E(G) = 0 otherwise
The resistance matrix (Klein and Randić 1993, and Bonchev *et al.* 1994) RM=RN(G)	see reference because the description of RM(G) is too long.
Cluj matrix(Diudea 1997, Diudea *et al.* 1997, Kiss *et al.* 1997, and Gutman 1997) Cj=Cj(G)	$[\mathrm{Cj}_u]_{ij} = \max\{\mathrm{N}_{i,p(m)} : m = 1, 2, \cdots\}$ where the $N_{i,p(m)}$ represent the number of vertices on each side of the path p_{ij}.

In order to extract structure information as much as possible, many methods focused their attention on describing atom and bond.

2.1 Matrices

In molecular graph, vertex represents atom and edge symbolizes bond. Thus, the molecular graphs can be easily expressed by matrices. Based on the matrix expression of the molecular graph, matrix polynomial, determinant, path, walk, and distance can be calculated (Berenike and Joachim 2001, Gutman *et al.* 2001, and Dayantis 1997). Matrix is the basis to compute other parameters. Some commonly used matrices expressing the molecular graphs are summarized in Table 1.

After getting the matrix of molecular graph, the characteristic determinants and spectra, which have important applications in molecular orbital theory (Graovac *et al.* 1977, Gutman and Polansky 1986, Trinajstic 1992, and Knop and Trinajstic 1980) and also are important sources of molecular descriptors (Hosoya 1971, 1988, 1990, Trinajstic 1988, and Hosoya and Murakami 1975), can be calculated. There are many references (Graovac *et al.* 1977, Gutman and Polansky 1986, Trinajstic 1988, 1992, Knop and

Table 2: Topological indices and their definitions*

Name of index (reference)	Definition
Autocorrelation descriptors (Moreau and Broto 1980)	$$T = (T_0, T_1, T_2) \quad T_n = g(i) * g(j),$$ where $g(i)$ and $g(j)$ are the contributions attributed to atoms i and j.
Balaban index J (Balaban 1982)	$$J = \Sigma(\bar{S}_i\bar{S}_j)^{-1/2} = q\Sigma(S_iS_j)^{-1/2},$$ where S_i and S_j mean the distance sums of the vertices V_i and V_j; $q = n_e/(\mu + 1)$, in which n_e is the number of edges and μ is the cycle number.
Bond flexibility index (Lieth *et al.* 1996)	$$\rho_{\mathrm{KB}} = \Sigma\Phi_i - \Phi + 1,$$ in which Φ_i corresponds to the fragment flexibilities; and Φ denotes the whole molecule flexibilities.
Centric indices (Balaban 1979, and Hu *et al.* 2003a)	$$B = \sum_i \delta_i^2 \quad C = 1/2(B - 2n + U),$$ where δ_i is the vertices at each step; and U is the Kronecker delta depending on the parity of the number of vertices n. $$CI = \sum_{i=1}^{n} BF_i, \text{ in which } BF_i = \sum_{j=i}^{n} I_j * V_j * D_{ij}$$ where I_j is the intrinsic state of atoms j, V_j is the vertex degree of atom j, and D_{ij} is the distance between atoms i and j.
Detour index (Lukovits 1996, Razinger 1997)	$$\omega = (1/2)\sum_i\sum_j(\Delta)_{ij},$$ where Δ_{ij} is an element in the detour matrix.
Edge connectivity index (Estrada 1995)	$$\varepsilon = \sum_r((\delta(e_i)(\delta(e_j)))_r^{-0.5},$$ where $\delta(e_i)$ and $\delta(e_j)$ are the degree of edges e_i and e_j.

Table 2 (continued): Topological indices and their definitions*

Name of index (reference)	Definition
Electropy index (Yee *et al.* 1977)	$\varepsilon = \log_2 \frac{N!}{\Pi N_i!}$, where N is the total number of atoms in the molecule. N_i means the number of one kind of atoms.
(E-state) index (Kier and Hall 1990)	$$S_i = I_i + \sum_j \Delta I_{ij},$$ where I_i is the intrinsic state value of atom i; ΔI_{ij} is the perturbation of I_j on I_i with the form as $\Delta I_{ij} = (I_i - I_j)/D_{ij}^2$.
Extended ECI (Estrada *et al.* 1998)	$${}^m\varepsilon_t(G) = \sum_s \prod_i [\delta(e_i)]_s^{-0.5},$$ where $\delta(e_i)$ is the degree of edge e_i.
Extended WI (Estrada *et al.* 1988)	$${}^kW = \sum_{i=k-1}^{l} i(i-1)(i-2)\cdots(i-(k-1))\eta_i,$$ in which η_i is the number of pairs of vertices at distance i.
Flexibility index (Kier 1989)	$$\Phi = {}^1\kappa_\alpha {}^2\kappa_\alpha / A,$$ where A is the number of atoms; ${}^1\kappa_\alpha$ and ${}^2\kappa_\alpha$ are the first and second order of Kappa index.
Fragment WI (Mekenyan *et al.* 1988)	$$W.\ EFTI\ (G) = \sum_{i=1}^{N} W.\ EFTI(v_i),$$ the index is to reflect the interaction between the excised fragment F and the remainder of the molecular graph $(G - F)$.
GSI (Gordon and Scantlebury 1964)	$$2N_2 = \sum_{i,j}^{n} (V_i + V_j) - 2,$$ where V_i and V_j are the vertex degrees of atoms i and j.

Table 2 (continued): Topological indices and their definitions*

Name of index (reference)	Definition
Harary index (Plavsic *et al.* 1993)	$H = (1/2)\sum_{i=1}^{N}\sum_{j=1}^{N}(D^r)_{ij}.$
Hosoya index (Hosoya 1971)	$Z(G) = \sum_{K=0}^{n/2} P(G, K),$ where $p(G, K)$ is the number of ways in which k edges of the graph may be chosen so that no two of them are adjacent.
Hyper-Wiener index (Klein *et al.* 1995)	$WW(G) = \frac{1}{2}\sum_{i<j}([D]_{ij}^2 + [D]_{ij}).$
Identification numbers (Randić 1977)	$ID = \frac{1}{2}[N + \sum_{i=1}^{T} W_i].$
Kappa index (Kier 1985, 1986)	${}^1\kappa = 2^1P\text{max}^1P\text{min}/({}^1P_i)^2,$ ${}^2\kappa = 2^2P\text{max}^2P\text{min}/({}^2P_i)^2,$ ${}^32\kappa = 4^3P\text{max}^3P\text{min}/({}^3P_i)^2,$ where ${}^1P_i, {}^2P_i$, and ${}^3P_i$ are the numbers of one, two, and three paths; ${}^1P_{\text{min}}$, and ${}^1P_{\text{max}}$ are the number of one-bond path of linear and complete graph; ${}^2P_{\text{min}}$, and ${}^2P_{\text{max}}$ are the number of two-bond paths of linear and star graph; ${}^3P_{\text{max}}$, and ${}^3P_{\text{max}}$ are the number of three-bond paths of linear and twin star graph.
Kirchoff index (number) (Klein and Randić 1993)	$Kf = \frac{1}{2}\sum_i\sum_j \Omega_{ij},$ where Ω_{ij} is the (i, j) element of resistance matrix.

Table 2 (continued): Topological indices and their definitions*

Name of index (reference)	Definition
Molecular connectivity index (Randić 1975, and Kier and Hall 1976)	$\chi = \sum (V_i V_j)^{-1/2}$, $^k\chi_p = \sum_{all\ edges} (V_i V_j \cdots V_k)^{-1/2}$, where V_i, V_j, and V_k are the vertex degrees of atoms i, j, and k.
Molecular topological index (Schultz 1989)	$MTI = \sum_{i=1}^{n} E_i$, in which E_i is the row matrix consisting of v(A+D) where v is the vertex degree.
Overall connectivity (Bonchev 1997)	$TC(G) = \sum_{k=1}^{K} A_k(G_k \in G) = \sum_{k=1}^{K}\sum_{i=1}^{N_k} a_i$.
Quasi-Wiener index (Mohar *et al.* 1993)	$W^*(G) = N \sum_{i=1}^{N-1} \frac{1}{Sp(L)_i}$, where $S_p(L)$ is the spectrum of the Laplace matrix.
Szeged index (Khadikar *et al.* 1995)	$S_z(G) = \sum_{e_{ij} \in E(G)} n_i n_j$.
Wiener index (Wiener 1947)	W(G)=1/2(sum(sum(d$_{ij}$)).
Zagred group indices (Gutman *et al.* 1975)	$M1 = \sum_{i=1}^{n} V_1^2 \quad M2 = \sum_{all\ edges} V_i V_j$, where V_i, and V_j are the vertex degrees of atoms i, and j.

*: The meaning of the symbols can be found in the reference listed.

Trinajstic 1980, Hosoya 1971, 1988, 1990, Hosoya and Murakami 1975, and Harray 1969) to discuss determinants and spectra. The commonly used characteristic determinants are: $Ch(A, G) = \det(xI - A)$ (Graovac *et al.*

1977, Gutman and Polansky 1986, Harary 1969, Cvetkovic *et al.* 1995), $Ch(D, G) = \det(xI - D)$ (Hosoya *et al.* 1973, Graham *et al.* 1977, and Graham and Lovasz 1978), $Ch(RD, G) = \det(xI - RD)$ (Diudea *et al.* 1997), and $Ch(L, G) = \det(xI - L)$ (Gutman *et al.* 1994, Trinajstic *et al.* 1994).

2.2 Topological index

There have been more than 400 kinds of topological indices available, since the birth of the first one. Topological index can be used to evaluate structural similarity and diversity. Its main role is to work as a numerical molecular descriptor in QSAR/QSPR model (Ivanciuc *et al.* 1999). Some important indices are listed in Table 2.

2.3 Atomic attribute

Nowadays, in addition to data on molecular connectivity, the information encoded by topological indices also includes the nature of atoms and the bond multiplicity. In topological description, another important aspect is to describe atoms and bonds, especially in weighted graph. Ivanciuc *et al.* has written a review (Ivanciuc and Balaban 1999) on the main schemes for computing vertex- and edge-weighted graph parameters, and the related structural descriptors. Ivanciuc *et al.* (1998) proposed to apply atomic electronegativity and covalent radius to obtain descriptors dealing with weighted graph.

The methods to describe nature of atoms and bond multiplicity first introduced some chemical parameters, such as atomic order (Z), relative eletronegativity (X), length of covalent radius (Y), atomic mass (A), atomic and adjacent hydrogen mass (AH), atomic polarity (P), atomic radius (R), and atomic eletronegativity (E). Based on these parameters, some topological invariants or indices were proposed (Nikolic *et al.* 1993, Balaban 1986). These descriptors were further applied to QSAR/QSPR researches (Medic *et al.* 1992, Balaban *et al.* 1990, 1992, Ivanciuc *et al.* 2000, Ivanciuc 2000, and Estrada 1997). The methods are summarized in Table 3 in detail.

Table 3: Schemes for describing weighted graph

Attribute Schemes	Definition for atoms and bonds	Reference
A	$V(A)_i = 1 - Ac/Ai$, $E(A)_{ij} = AcAc/B_{oij}A_iA_j$, where Ac is the atomic mass of carbon atom; A_i, and A_j are the atomic masses of atom i and j. B_{oij} is the topological bond order of the edge between atom i and j.	Ivanciuc and Balaban 1999
AH	$V(AH)_i = 1 - Ac/(A_i + NoH_iA_H)$, $E(AH)_{ij} = AcAc/B_{oij}(A_i + NoH_iA_H)(A_j + NoH_jA_H)$, where NoH_i is the number of hydrogen atoms bonded to the heavy atom i.	Ivanciuc and Balaban 1999
Z	$V(Z)_i = 1 - Zc/Z_i = 1 - 6/Z_i$, $E(Z)_{ij} = ZcZc/B_{oij}Z_iZ_j$, where Zc is the atomic order of carbon atom; Z_i, and Z_j are the atomic orders of atom i and j. B_{oij} is the topological bond order of the edge between atom i and j.	Ivanciuc *et al.* 1998, and Barysz *et al.* 1983
X	$V(X)_i = 1 - 1/X_i$, $E(X)_{ij} = 1/B_{oij}X_iX_j$, where X_i, and X_j are the relative eletronegativity of atom i and j. B_{oij} is the topological bond order of the edge between atom i and j.	Balaban 1986, and Ivanciuc *et al.* 1998
Y	$V(Y)_i = 1 - 1/Y_i$, $E(Y)_{ij} = 1/B_{oij}Y_iY_j$, where Y_i and Y_j are the length of covalent radius of atom i and j. B_{oij} is the topological bond order of the edge between atom i and j.	Sanderson 1983, Ivanciuc 1999
P	$V(P)_i = 1 - \alpha_C/\alpha_i = 1 - 1.76/\alpha_i$, $E(P)_{ij} = \alpha_C\alpha_C/B_{oij}\alpha_i\alpha_j$, where α_C is the atomic polarity of carbon atom; α_i, and α_j are the atomic polarity of atom i and j. B_{oij} is the topological bond order of the edge between atom i and j.	Ivanciuc *et al.* 1998, and Nagle 1990

Table 3 (continued): Schemes for describing weighted graph

Attribute Schemes	Definition for atoms and bonds	Reference
R	$V(R)_i = 1 - r_C/r_i = 1 - 1.21/r_i$, $E(P)_{ij} = r_C r_C / B_{oij} r_i r_j$, where r_C is the atomic radius of carbon atom; r_i, and r_j are the atomic radius of atom i and j. B_{oij} is the topological bond order of the edge between atom i and j.	Ivanciuc *et al.* 1998, and Nagle 1990

2.4 An example

In order to make it easier for readers to understand how to obtain the numerical molecular descriptor, an example is given on a saturated hydrocarbon named lipC4, together with its hydrogen-depressed graph as shown in Figure 1. The numbers, say 1, 2, $\cdots$, 7, are the labels of the atoms in the graph.

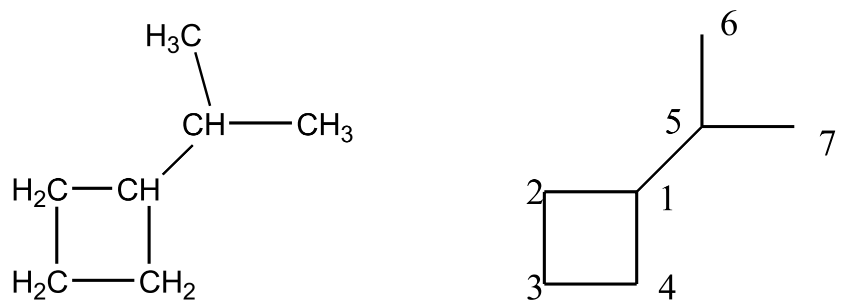

Figure 1. Molecular structure and the corresponding topological graph (hydrogen-depressed) of 1ipC4

With the help of the labeled hydrogen-depressed graph, one could easily get its adjacent, distance and detour matrices (longest path) of the molecule according to their definitions (see Table 1). The labeled numbers in the graph correspond to the number of the row or the column in the matrix. Take the adjacent matrix as an example, if $i \neq j$ and $e_{ij} \in E(G)$, then $[A]_{ij} = 1$; and if $i = j$ or $e_{ij} \notin E(G)$, then $[A]_{ij} = 0$. In this way, the adjacent matrix can be easily obtained as shown in Table 4. Distance matrix can be achieved in the same way. If $i \neq j$, then $[D]_{ij} = d_{ij}$, that is, the distance between these two vertexes, and if $i = j$, then $[D]_{ij} = 0$. Similarly, the detour matrix (its elements are the longest distances between pairs of atoms) is also shown in the Table 4. From the Distance matrix in Table 4, we can see that the element at column 5 and row 3 is 3, which means that the distance between the vertex 5 and vertex 3 is 3.

Table 4: The adjacent, distance and detour matrices of molecule 1ipC4

	Adjacent matrix							Distance matrix							Detour matrix						
	1	2	3	4	5	6	7	1	2	3	4	5	6	7	1	2	3	4	5	6	7
1	0	1	0	1	1	0	0	0	1	2	1	1	2	2	0	3	2	3	1	2	2
2	1	0	1	0	0	0	0	1	0	1	2	2	3	3	3	0	3	2	4	5	5
3	0	1	0	1	0	0	0	2	1	0	1	3	4	4	2	3	0	3	3	4	4
4	1	0	1	0	0	0	0	1	2	1	0	2	3	3	3	2	3	0	4	5	5
5	1	0	0	0	0	1	1	1	2	3	2	0	1	1	1	4	3	4	0	1	1
6	0	0	0	0	1	0	0	2	3	4	3	1	0	2	2	5	4	5	1	0	2
7	0	0	0	0	1	0	0	2	3	4	3	1	2	0	2	5	4	5	1	2	0

With the adjacent and distance matrices at hand, some topological indices (see Table 2) can be obtained directly from them, such as W (Wiener index). One can simply sum up all the distances between the vertexes to get it (see Table 5). However, some others need some complicated operations. Table 5 lists some examples for the molecule 1ipC4 shown in Figure 1.

Here we will take molecular connectivity index as an example to illustrate the calculation procedure. The molecular connectivity index is to describe the molecular connectivity (see its definition in Table 2). For instance, if we want to get χ^1, we need to get the vertex degree (V_i), which is the sum of i row of A matrix for atom i, for every vertex in the molecular graph. From Figure 1, one can easily obtain such information, that is, $V_1 = 3, V_2 = 2, V_3 = 2, V_4 = 2, V_5 = 3, V_6 = 1$, and $V_7 = 1$. Then, we can use the equation, say $\chi^1 = \sum (V_i V_j)^{-1/2}$, to do the calculation first for every two adjacent vertexes and then sum them up to get χ^1 as shown in Table 5. In general, the topological indices offer a simple way of coding molecular structure information into numerical values.

Table 5: Some topological indices of molecule 1ipC4

Index name	Definition	Value
W	$W(G) = 1/2(\sum(\sum(d_{ij})))$	44
	$\omega = (1/2)\sum_i \sum_j (\Delta_{ij})$	
Detour Index		64
χ^0	$\chi^0 = \sum(V_i)^{-1/2}$	5.2760
χ^1	$\chi^1 = \sum(V_iV_j)^{-1/2}$	3.3045
χ^2	$\chi^2 = \sum(V_iV_jV_k)^{-1/2}$	2.9350
Kappa^1	${}^1\kappa = 2^1P\text{max}^1P\text{min}/({}^1P_i)^2$	5.1429
Kappa^2	${}^2\kappa = 2^2P\text{max}^2P\text{min}/({}^2P_i)^2$	1.8519
Kappa^3	${}^3\kappa = 4^3P\text{max}^3P\text{min}/({}^3P_i)^2$	0.9600

3. Some Directions in Developing Topological Index

Although the topological indices have wide applications in QSAR/QSPR, there exit some shortcomings:

a. Compared with other parameters, the physical-chemical meaning of topological index is not explicit (Devillers *et al.* 1997).

b. The degree of redundancy and degeneracy of certain topological indices can be very high. In that case, it is impossible to interpret the QSAR and QSPR models derived from these descriptors (Devillers 1999), which is worthy to study the problem under these kinds of situations and the topological indices should be employed only in contexts for which they are suitable.

c. Topological index usually neglects the stereochemical information or the three-dimensional structure of the molecule (Ivanciuc *et al.* 1999).

To deal with these shortcomings, different methods are proposed to improve the topological index. The main directions of topological index are mainly following points:

(1) Description of local information

Recent enrichments in these areas include topological indices for molecular fragments, some stereochemical features and electronic parameters associated with various atoms. The introduction of local information is helpful to explain the physical or chemical meaning of topological index. Local descriptors can be geometric, steric, hydrophobic, hydrophilic constants, or atomic electron density. Estrada thinks that the topological index is successful if the index has direct structural or physical meaning and at same time obtains similar results of original QSAR/QSPR model (Estrada 1999). A developing direction is to include not only topological but geometric characteristics to deal with three-dimensional space of stereochemistry. Another direction is to utilize atomic property to develop new index or improve the original index.

(2) Studies on inter-correlation of topological index

Estrada thinks the resolution of practical problems should include as many descriptors as possible and the built QSAR/QSPR model should select as few descriptors as possible (Estrada 1999). The problem of reducing number of variables makes the study on correlation between topological indices important.

In QSAR research, a very important problem is how to reduce the number of variables to improve the stability of the model. On one side, the information of many topological indices is duplicated. On another side, how to select the required variables from the lots of topological indices is still unsolved. The solution of the problems requires the study on correlation between variables and there are some primary studies (Motoc and Balaban 1981, Motoc *et al.* 1982, Plavsic *et al.* 1996, 2000, Randić 2001, Rücker and Rücker 1994, and Chan *et al.* 1998).

(3) Variable index

The simple index should be modified if the molecules contain heteroatoms. The variable index is regarded as a novel way to describe the heteroatoms. The optimal value is computed through the regression procedure. The variable index is a flexible function, which makes the standard error of regression minimum and then find the optimal number (Randić and Pompe 2001b). There are many prior studies

(Randić and Basak 2001, Randić and Pompe 2001a, Randić *et al.* 2001) on the variable index and the authors also proposed one new variable index (Hu *et al.* 2003b).

Acknowledgement

The authors appreciate valuable comments from Prof. Min-Te Chao. The first two authors were financially supported by National Nature Foundation Committee (NNFC) of P. R. China (No.20235020, 20175036) and they appreciate the hospitality of Hong Kong Baptist University, when the authors attend "Workshop on Data Mining in Traditional Chinese Medicines". The third author was partially supported by the Statistics Research and Consultancy Centre, Hong Kong Baptist University.

References

Agrawal, V. K. (2001). A novel PI index and its applications to QSPR/QSAR studies. *J. Chem. Inf. Comput. Sci.*, **41**, 934-949.

Amic, D. and Trinajstic, N. (1995). On the detour matrix. *Croat. Chem. Acta.*, **68**, 53-62.

Balaban, A. T. (1979). Chemical graphs. Part 35. Five new topological indices for the branching of tree-like graphs. *Theoret. Chim. Acta (Berl.)*, **5**, 239-261.

Balaban, A. T. (1982). Highly discriminating distance-based topological index. *Chem. Phys. Lett.*, **89**, 399-404.

Balaban, A. T. (1986). Chemical graphs. Part 48. Topological index *J* for heteroatom-containing molecules taking into account periodicities of element properties. *MATCH*, **21**, 115-122.

Balaban, A. T. (1995). Chemical graphs: Looking back and glimpsing ahead. *J. Chem. Inf. Comput. Sci.*, **35**, 339-350.

Balaban, A. T. (1998). Real-number vertex invariants and Schultz-type indices based on eigenvectors of adjacency and distance matrices. *J. Chem. Inf. Comput. Sci.*, **38**, 1038-1047.

Balaban, A. T., Catana, C. D. and Niculescu, D. I. (1990). Applications of weighted topological index J for QSAR of Carcinogenesis inhibitors (Retinoic acid derivatives). *Rev. Roum. Chin.*, **35**, 997-1003.

Balaban, A. T., Kier, L. B. and Joshi, N. (1992). Correlations between chemical structure and normal boiling points of acyclic ethers, peroxides, acetals and their sulfur analogues. *J. Chem. Inf. Comput. Sci.*, **32**, 237-244.

Barysz, M., Jashari, G., Lall, R. S., Srivastava, V. K. and Trinajstic, N. (1983). On the distance matrix of molecules containing heteroatoms. In *Chemical Application of Topology and Graph Theory.* (King, P. eds.), Elsevier, Amsterdam. p.222-227.

Basak, S. C. and Grunwald, G. D. (1994). Molecualr similarity and risk assessment: analog selection and property estimation using graph invariants. *SAR QSAR Environ. Res.*, **2**, 289-307.

Basak, S. C., Bertelsen, S. and Grunwald, G. D. (1994). Application of graph theoretic parameters in quantifying molecular similarity and structure-activity relationships. *J. Chem. Inf. Comput. Sci.*, **34**, 270-276.

Basak, S. C., Gute, B. D. and Grunwald, G. D. (1999). A hierarchical approach to the development of QSAR models using topological, geometrical and quantum chemical parameters. In *Topological Indices and Related Descriptors in QSAR and QSPR.* (Devillers, J. and Balaban, A.T. eds.), Gordon and Breach Science Publishers. p 692.

Berenike, M. and Joachim, O. R. (2001). Shape of self-avoiding walks in two dimensions, *Macromolecules*, **34**, 5723-5724.

Bonchev, D. (1997). Novel indices for the topological complexity of molecules. *SAR. QSAR. Environ. Res.*, **7**, 23-43.

Bonchev, D. (2000). Overall connectivities/topological complexities: a new powerful tool for QSPR/QSAR. *J. Chem. Inf. Comput. Sci.*, **40**, 934-941.

Bonchev, D. (2001). The overall Wiener index-a new tool for characterization of molecular topology. *J. Chem. Inf. Comput. Sci.*, **41**, 582-592.

Bonchev, D. and Tinajstic, N. (1977). Information theory, distance matrix, and molecular branching. *J. Chem. Phys.*, **67**, 4517-4533.

Bonchev, D., Balaban, A. T., Liu, X. and Klein, D. J. (1994). Molecualr cyclicity and centricity of polycyclic graphs. I. Cyclicity based on resistance distances or reciprocal distances. *Int. J. Quantum. Chem.*, **50**, 1-20.

Bono, L., Istvan, L., Sonja, N. and Trinajstic, N. (2001). Distance-related indexes in the quantitative structure-property relationship modeling. *J. Chem. Inf. Comput. Sci.*, **41**, 527-535.

Cao, C. Z. and Yuan, H. (2001). Topological indices based on vertex, distance, and ring: on the boiling points of paraffins and cycloalkanes. *J. Chem. Inf. Comput. Sci.*, **41**, 867-877.

Chan, O., Gutman, I., Lam, T. K. and Merris, R. (1998). Algebraic connections between topological indices. *J. Chem. Inf. Comput. Sci.*, **38**, 62-65.

Christian, D. P., Klein, M. K., Christiane, S., Silke, L., Stefanie, H., Dieter, H., Klaus, M. and A. J. Hopfinger. (1999). Synthesis, pharmacological and biophysical characterization, and membrane-interaction QSAR analysis of cationic amphiphilic model compounds. *J.*, **42**, 3874-3888.

Cvetkovic, D. M., Doob, M. and Sachs, H. (1995). *Spectra of Graphs. Theory and Applications*, 3rd edition. Johann Ambrosius Barth Verlag, Heidelberg.

David, A. C. (2000). QSPR correlation and predictions of GC retention indexes for methyl-branched hydrocarbons produced by insects. *Analytical Chemistry*, **72**, 101-109.

David, F. V. (2000). Structural characteristics of human P450s involved in drug metabolism: QSARs and lipophilicity profiles. *Toxicology*, **144**, 197-203.

Dayantis, J. (1997). Monte Carlo simulations of free and confined walks in reflecting statistics. *J. Chem. Inf. Comput. Sci.*, **37**, 501-509.

Devillers, J. (1999). No-free-lunch molecular descriptors in QSAR and QSPR. In *Topological Indices and Related Descriptors in QSAR and QSPR*. (Devillers, J. and Balaban, A. T. eds.) Gordon and Breach Science Publishers, P1-17.

Devillers, J. and Balaban, A. T. (1999). *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon and Breach Science Publishers.

Devillers, J., Domine, D., Guillon, C., Bintein, S. and Karcher, W. (1997). Prediction of partition coefficients (logP_{oct}) using autocorrelation descriptors. *SAR QSAR Environ. Res.*, **7**, 151-172.

Diudea, M. V. (1997). Cluj matrix invariants. *J. Chem. Inf. Comput. Sci.*, **37**, 300-305.

Diudea, M. V. (2000). *QSPR/QSAR Studies by Milecular Descriptors.* Nova Science Publishers.

Diudea, M. V. and Ivanciuc, O. (1995). *Molecular topology.* Comprex, Cluj, Romania.

Diudea, M. V., Horvath, D. and Graovac, A. (1995). Molecular topology. 15.3D distance matrices and related topological indices. *J. Chem. Inf. Comput. Sci.*, **35**, 129-135.

Diudea, M. V., Katona, G., Lukovits, I. and Trinajstic, N. (1998). Detour and Cluj-detour indices. *Croat. Chem. Acta.*, **71**, 459-471.

Diudea, M. V., Parv, B. and Gutman. I. (1997). Detour-Cluj matrix and derived invariants. *J. Chem. Inf. Comput. Sci.*, **37**, 1101-1108.

Erovnik, J. (1999). Szeged index of symmetric graphs. *J. Chem. Inf. Comput. Sci.*, **39**, 77-80.

Estrada, E. (1995). Edge adjacency relationships and a novel topological index related to molecular volum. *J. Chem. Inf. Comput. Sci.*, **35**, 31-33.

Estrada, E. (1996). Spectra moments of the edge adjacency matrix in molecular gaphs. 1. Definition and applications to the prediction of physical properties of alkanes. *J. Chem. Inf. Comput. Sci.*, **36**, 844-849.

Estrada, E. (1997). Spectral moments of the edge-adjacency matrix of molecular graphs. 2. Molecules containing heteroatoms and QSAR applications, *J. Chem. Inf. Comput. Sci.*, **37**, 320-328.

Estrada, E. (1999). Novel strategies in the search of topological indices. In *Topological Indices and Related Descriptors in QSAR and QSPR.* (Devillers, J. and Balaban, A. T. eds.) Gordon and Breach Science Publishers, P404-454.

Estrada, E. and Molina, E. (2001). 3D connectivity indices in QSPR/QSAR studies. *J. Chem. Inf. Comput. Sci.*, **41**, 791-797.

Estrada, E. and Ramirez, A. (1996). Edge adjacency relationships and molecular topographic descriptors. Definition and QSAR applications. *J. Chem. Inf. Comput. Sci.*, **36**, 837-843.

Estrada, E., Guevara, N. and Gutman, I. (1998). Extension of edge connectivity index. Relationships to line graph indices and QSPR applications. *J. Chem. Inf. Comput. Sci.*, **38**, 428-431.

Estrada, E., Ivanciuc, O., Gutman, I., Gutierrez, A. and Rodriguez, L. (1988). Extended Wiener indices. A new set of descriptors for quantitative structure-propery studies. *New J. Chem.*, **22**, 819-822.

Fujita, T. and Timmerman, H. (1995). *QSAR and Drug Design: New Developments and Applications*. Elsevier Science.

Gao, H., John, A., Katzenellenbogen, R. G. and Hansch, C. (1999). Comparative QSAR analysis of estrogen receptor ligands. *Chemical Reviews*, **99**, 723-744.

Geiss, K. T. and Frazier, J. M. (2001). QSAR modeling of oxidative stress in vitro following hepatocyte exposures to halogenated methanes. *Toxicology in Vitro*, **15**, 557 -563.

Gordon, M. and Scantlebury, G. R. (1964). Nonrandom polycondensation: statistical theory of the substitution effect. *Trans. Faraday Soc.*, **60**, 604-621.

Graham, R. L. and Lovasz, L. (1978). Distance matrix polynomials of trees. *Adv. Math.*, **29**, 60-88.

Graham, R. L., Hoffman, A. J. and Hosoya, H. (1977). On the distance matrix of a directed graph. *J. Graph Theory*, **1**, 85-88.

Graovac, A., Gutman, I. and Trinajstic, N. (1977). *Topological Approach to the Chemistry of Conjugated Moleculaes*. Springer, Berlin.

Gutman, I. (1997). Detour-Cluj matrix and derived invariants. *J. Chem. Inf. Comput. Sci.*, **37**, 1101-1108.

Gutman, I. and Polansky, O. E. (1986). *Mathematical Concepts in Organic Chemistry*, Springer, Berlin.

Gutman, I., Bonchev, D. and Rouvray, D. H. (1991). *Chemical Graph Theory: Introduction and Fundamentals, Vol. 1.* Gordon & Breach Publishing Group.

Gutman, I., Lee, S. L., Chu, C, H. and Luo, Y. L. (1994). Chemical applications of the Laplacian spectrum of molecular graphs: studies of the Wiener number. *Ind. J. Chem.*, **33A**, 603-608.

Gutman, I., Rücker, C. and Rücker, G. (2001). On walks in molecular graphs. *J. Chem. Inf. Comput. Sci.*, **41**, 739-745.

Gutman, I., Rusici, B., Trajstic, N. and Wilcox, C. F. (1975). Graph theory and molecular orbits. XII. Acyclic polyenes. *J. Chem. Phys.*, **62**, 3399-3409.

Hansch, C., Kurup, A., Garg, A. and Gao, H. (2001). Chem-Bioinformatics and QSAR: a review of QSAR lacking positive hydrophobic terms, *Chemical Reviews*, **101**, 619-672.

Harary, F. (1969). *Graph Theory.* Addison-Wesley, Reading, MA.

Hosoya, H. (1971). A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. *Bull. Chem. Soc. Japan.*, **44**, 2332-2339.

Hosoya, H. (1988). On some counting polynomials in chemistry. *Discrete Appl. Math.*, **19**, 239-257.

Hosoya, H. (1990). Some recent advances in counting polynomials in chemical graph theory. In *Computational chemical graph theory* (Rouvray, D.H. eds.). Nova Science Publishers, New York, pp.105-126.

Hosoya, H. and Murakami, M. (1975). Topological index as applied to π-electronic systems. II. Topological bond order. *Bull. Chem. Soc. Japan.*, **48**, 3512-3517.

Hosoya, H., Murakami, M. and Gotoh, M. (1973). Distance polynomial and characterization of a graph. *Natl. Sci. Rept. Ochanomizu Univ.*, **24**, 27-34.

Hu, Q. N., Liang, Y. Z. and Ren, F. L. (2003a) Molecular graph center, a novel approach to locate the center of a molecule and a new centric index. *J. Mol. Struct. THEOCHEM.* (in press).

Hu, Q. N., Liang, Y. Z., Wang, Y. L., Xu, C. J., Zeng, Z. D., Fang, K. T., Peng, X. L., Yin, H. (2003b). External factor variable connectivity index. *J. Chem. Inf. Comput. Sci.*, **43**, 773-778.

Huang, X. Q., Xu, L. S, Luo, X. M., Fan, K. N., Ji, R. Y., Pei, G., Chen, K. X. and Jiang H. L. (2002). Elucidating the inhibiting mode of AHPBA derivatives against HIV-1 protease and building predictive 3D-QSAR models. *Journal of Medicinal Chemistry*, **45**, 333-343.

Ivanciuc, O. (1989). Design on topological indices. Part 1. Definition of a vertex topological index in the cases of 4-tree. *Rev. Roum. Chim.*, **34**, 1361-1368.

Ivanciuc, O. (1993). Chemical graph polynomials. Part 3. The Laplacian polynomial of molecular graphs. *Rev. Roum. Chim.*, **38**, 1499-1508.

Ivanciuc, O. (1999). Design of topological indices. Part 12. parameters for vertex- and edge-weighted molecular graphs. *Rev. Roum. Chim.*

Ivanciuc, O. (2000). QSAR comparative study of Wiener descriptors for weighted molecular graphs. *J. Chem. Inf. Comput. Sci.*, **40**, 1412-1422.

Ivanciuc, O. and Balaban, A. T. (1994). Design of topological indices. Part 8. Path matrices and derived molecular graph invariants. *MATCH*, **30**, 141-152.

Ivanciuc, O. and Balaban, A. T. (1999). Vertex-and edge-weighted molecular graphs and derived structural descriptors. In *Topological Indices and Related Descriptors in QSAR and QSPR.* (James Devillers and Alexandru T. Balaban. eds.) Gordon and Breach Science Publishers, p171-220.

Ivanciuc, O., Balaban, T. S. and Balaban, A. T. (1993). Design of topological indices. Part 4. Reciprocal distance matrix, related local vertex invariants and topological indices. *J. Math. Chem.*, **12**, 309-318.

Ivanciuc, O., Ivanciuc, T. and Balaban, A. T. (1998). Design of topological indices. Part 10. Parameters based on electronegativity and covalent radius for the computation of molecular graph descriptors for heteroatom-containing molecules. *J. Chem. Inf. Comput. Sci.*, **38**, 395-401.

Ivanciuc, O., Ivanciuc, T. and Balaban, A. T. (1999). The graph description of chemical structures. In *Topological Indices and Related Descriptors in QSAR and QSPR*. (James Devillers and A. T. Balaban. eds.) Gordon and Breach Science Publishers, P60-61.

Ivanciuc, O., Ivanciuc, T., Daniel, C. B. and Balaban, A. T. (2000). Comparison of weighting schemes for molecular graph descriptors: Application in quantitative structure-retention relationship models for Alkylphenols in gas-liquid chromatography. *J. Chem. Inf. Comput. Sci.*, **40**, 732-743.

Karcher, W. and Deviller, J. (1990). *Practical Applications of Quantitative Structure-Activity Relationships (QSAR) in Environmental Chemistry and Toxicology.* Kluwer Academic Publishers.

Karmarkar, S. and Khadikar, P. V. (2000). A comparative QSAR study using Wiener, Szeged, and molecular Connectivity Indices. *J. Chem. Inf. Comput. Sci.*, **40**, 57-62.

Katritzky, A. R. (1996). Quantum-chemical descriptors in QSAR/QSPR studies, *Chemical Reviews*, **96**, 1027-1044.

Katritzky, A. R. and Gordeeva E. V. (1993). Traditional topological indices vs electronic, geometrical and combined molecular descriptors in QSAR/QSPR research. *J. Chem. Inf. Comput. Sci.*, **33**, 835-857.

Khadikar, P. V., Deshphande, N. V., Kale, P. P., Dobrynin, A. A., Gutman, I. (1995). The Szeged index and an analogy with the Wiener index. *J. Chem. Inf. Comput. Sci.*, **35**, 547-550.

Kier, L. B. (1985). A shape index from chemical graphs. *Quant. Struct. Act. Relat.*, **4**, 109-116.

Kier, L. B. (1986). Shape indexes of orders one and three from molecular graph. *Quant. Struct. Act. Relat.*, **5**, 1-7.

Kier, L. B. (1989). An index of molecular flexibility from kappa shape attribute. *Quant. Struct. Act. Relat.*, **8**, 221-224.

Kier, L. B. and Hall, L. H. (1976). *Molecular Connectivity in Chemistry and Drug Research.* Academic Press, New York, p 257.

Kier, L. B. and Hall, L. H. (1990). An electrotopological state index for atoms in molecules. *Pharm. Res.*, **7**, 801-807.

Kier, L.B. and Hall, L.H. (1999). *Molecular structure description: the electrotopological state*. Academic press.

King, R. B. (1992). *Applications of Graph Theory and Topology in Inorganic Cluster and Coordination Chemistry*. CRC Press, LLC.

King, R. B. and Rouvray, D. H. (1987). *Graph Theory and Topology in Chemistry*. Elsevier Science.

Kiss, A. A., Katona, G. and Diudea, M. V. (1997). Szeged and Cluj matrices within the matrix operator $W_{(M1, M2, M3)}$. *Coll. Sci. Pap. Fac. Sci. Kragujevac.*, **19**, 95-107.

Klein, D. J. and Randić, M. (1993). Resistance distance. *J. Math. Chem.*, **12**, 81-95.

Klein, D. J., Lukovits, I. and Gutman, I. (1995). On the definition of the hyper-Wiener index for cycle-containing structures. *J. Chem. Inf. Comput. Sci.*, **35**, 50-52.

Knop, J. V. and Trinajstic, N. (1980). Chemical graph theory. II. On the gaph theoretical polynomials of conjugated structures. *Int. J. Quantum. Chem.: Quantum Chem. Symp.*, **14**, 503-520.

Koh, L. L., Kon, O. L., Loh, K. W., Long, Y. C., Ranford, J. D., Tan, L. C. and Tjan, Y. Y. (1998). Complexes of salicylaldehyde acylhydrazones: cytotoxicity, QSAR and crystal structure of the sterically hindered t-butyl dimmer. *Journal of Inorganic Biochemistry*, **72**, 155-162.

Kubinyi, H., Folkers, G. and Martin, Y.C. (1998). *3D QSAR in Drug Design: Ligand-Protein Interactions and Molecular Similarity*, Vol. 2. Kluwer Academic Publishers.

Kurup, A., Garg, R. and Hansch, C. (2001). Comparative QSAR study of Tyrosine Kinase inhibitors, *Chemical Reviews*, **101**, 2573-2600.

Li, X. H., Yu, Q. S. and Zhu, L. G. (2000). A novel quantum-topology index. *J. Chem. Inf. Comput. Sci.*, **40**, 399-402.

Lieth, C. W., Stumpf, N. K. and Prior, U. (1996). A bond flexibility index derived from the constitution of molecules. *J. Chem. Inf. Comput. Sci.*, **36**, 711-716.

Liu, S. S., Yin, C. S., Li, Z. L. and Cai, S. X. (2001). QSAR study of steroid benchmark and dipeptides based on MEDV-13. *J. Chem. Inf. Comput. Sci.*, **41**, 321-329.

Lukovits, I. (1996). The detour index. *Croat. Chem. Acta.*, **69**, 873-883.

Lukovits, I. (2000). A compact form of the adjacency matrix,. *J. Chem. Inf. Comput. Sci.*, **40**, 1147-1150.

Madan, A. K. (1997). Eccentric connectivity index: a novel highly discriminating topological descriptor for structure-property and structure-activity studies. *J. Chem. Inf. Comput. Sci.*, **37**, 273-282.

Madan, A. K. (1999). Superpendentic index: a novel topological descriptor for predicting biological activity. *J. Chem. Inf. Comput. Sci.* **39**, 272-277.

Medic, S. M., Nikolic, S., Matijevic, S. J. (1992). A QSPR study of 3-(Phthalimidoalkyl)-pyrazolin-5-one. *Acta. Pharm.*, **42**, 153-167.

Mekenyan, O., Bonchev, D. and Balaban, A. T. (1988). Topological indices for molecular fragments and new graph invariants. *J. Math. Chem.*, **2**, 347-375.

Miguel, M. S., Facundo, P. G., Ricardo, N. M., Salabert, M. S., Francisco, J. G. M., Rosa, A., Cercos, P. and Teresa, M. G. (2001). QSAR analysis of hypoglycemic agents using the topological indices. *J. Chem. Inf. Comput. Sci.*, **41**, 1345-1354.

Mihalic, Z. (1997). The detour matrix in chemistry. *J. Chem. Inf. Comput. Sci.*, **37**, 631-638.

Mihalic, Z., Veljan, D., Nikolic, S., Plavsic, D. and Trinajstic, N. (1992). The distance matrix in chemistry. *J. Math. Chem.*, **11**, 223-258.

Mohar, B. (1989). Laplacian matrices of graphs. *Stud. Phys. Theor. Chem.*, **63**, 1-8.

Mohar, B., Babic, D. and Trinajstic, N. (1993). A novel definition of the Wiener index for trees. *J. Chem. Inf. Comput. Sci.*, **33**, 153-154.

Moreau, G. and Broto, P. (1980). The autocorrelation of a topological structure: a new molecular descriptor. *Nouv. J. Chim.*, **4**, 359-360.

Motoc, I. and Balaban, A. T. (1981). Topological indices. Intercorrelations, physical meaning, correlational ability. *Rev. Roum. Chim.*, **26**, 593-600.

Motoc, I., Balaban, A. T., Mekenyan, O. and Bonchev, D. (1982). Topological indices: inter-relations and composition. *MATCH*, **13**, 369-404.

Nagle, J. K. (1990). Atomic polarizability and electronegativity. *J. Am. Chem. Soc.*, **112**, 4741-4747.

Nikolic, S., Trajstic, N., Mihalic,Z. (1993). Molecular topological index: an extension to heterosystems. *J. Math. Chem.*, **12**, 251-264.

Plavsic, D., Nella, L. and Katica, S. B. (2000). On the relation between W'/W Index, hyper-Wiener index, and Wiener number. *J. Chem. Inf. Comput. Sci.*, **40**, 516-519.

Plavsic, D., Nikolic, S., Trinajstic, N. and Mihalic, Z. (1993). On the Harary index for the characterization of chemical graphs. *J. Math. Chem.*, **12**, 235-250.

Plavsic, D., Soskic, M., Landeka, I., Gutman, I. and Graovac, (1996). A. On the relation between the path numbers ^{1}Z,^{2}Z and the Hosoya Z index. *J. Chem. Inf. Comput. Sci.*, **36**, 1118-1122.

Pogliani, L. (2000). From molecular connectivity indices to semiempirical connectivity terms: recent trends in graph theoretical descriptors. *Chemical Reviews*, **100**, 3827-3858.

Rambon, C, D., Girones, E., Besalu, D. and Robert, L. A. (2000). M*olecular Quantum Similarity in QSAR and Drug Design.* Springer-Verlag New York, incorporated.

Randić, M. (1975). On characterization of molecular branching. *J. Am. Chem. Soc.*, **97**, 6609-6013.

Randić, M. (1977). On canonical numbering of atoms in a molecule and graph isomorphism. *J. Chem. Inf. Comput. Sci.*, **17**, 171-180.

Randić, M. (1992). Similarity based on extended basis descriptors. *J. Chem. Inf. Comput. Sci.*, **32**, 686-692.

Randić, M. (2001). Novel shape descriptors for molecular graphs. *J. Chem. Inf. Comput. Sci.*, **41**, 607-613.

Randić, M. and Basak, S. C. (2001). On use of the variable connectivity index $^1\chi^f$ in QSAR: toxicity of aliphatic ethers. *J. Chem. Inf. Comput. Sci.*, **41**, 614-618.

Randić, M. and Pompe, M. (2001a). The variable connectivity index $^1\chi^f$ versus the traditional molecular descriptors: a comparative study of $^1\chi^f$ against descriptors of CODESSA. *J. Chem. Inf. Comput. Sci.*, **41**, 631-638.

Randić, M. and Pompe, M. (2001b). The variable molecular descriptors based on distance related matrices. *J. Chem. Inf. Comput. Sci.*, **41**, 575-581.

Randić, M. and Zupan, J. (2001). On interpretation of well-known topological indices. *J. Chem. Inf. Comput. Sci.*, **41**, 550-560.

Randić, M., Plavi, D. and Ler, N. (2001). Variable connectivity index for cycle-containing structures. *J. Chem. Inf. Comput. Sci.*, **41**, 657-662.

Razinger, M. (1997). On calculation of the Detour index, *J. Chem. Inf. Comput. Sci.*, **37**, 283-286.

Rouvary, D. H. (1986). The role of the topological distance matrix in chemistry. In *Mathmatics and computational concepts in Chemistry.* (Trinajstic, N. eds.) Ellis Horwood, New York, pp295-306.

Rücker, C. (1998). Symmetry-aided computation of the Detour matrix and the Detour index. *J. Chem. Inf. Comput. Sci.*, **38**, 710-714.

Rücker, C. and Rücker, G. (1994). Mathematical relation between extended connectivity and eigenvector coefficients. *J. Chem. Inf. Comput. Sci.*, **34**, 534-538.

Rücker, G. and Rücker, C. (1999). On the topological indices, boiling points, and cycloalkanes. *J. Chem. Inf. Comput. Sci.*, **39**, 788-802.

Sanderson, R. T. (1983). *Polar Covalence.* Academic Press, New York, p.41.

Schultz, H. P. (1989). Topoloical organic chemistry. 1.Graph theory and topological indices of alkanes. *J. Chem. Inf. Comput. Sci.*, **29**, 227-228.

Schultz, H. P. (2000). Topological organic chemistry. 13. Transformation of graph adjacency matrixes to distance matrixes. *J. Chem. Inf. Comput. Sci.*, **40**, 1158-1159.

Shapiro, S., Rivero, N. and Carrasco, R. (1998). A QSAR study of quinolones based on electrotopological state index for atoms. *Quant. Struct. Act. Relat.*, **17**, 327-337.

Simona, T., Walter, S., Ludovic, K. and Zeno Simon. (2000). A review of QSAR for dye affinity for cellulose fibres. *Dyes and Pigments*, **47**, 5-16.

Suresh, B. M. and Hansch, C. (2001). Comparative QSAR studies on Bibenzimidazoles and Terbenzimidazoles inhibiting topoisomerase. *Bioorganic & Medicinal Chemistry*, **9**, 2885 -2893.

Trinajstic, N. (1988). The characteristic polynomial of a chemical graph. *J. Math. Chem.*, **2**, 197-215.

Trinajstic, N. (1992). *Chemical Graph Theory*, 2nd edition. CRC Press, Boca Raton, p 322.

Trinajstic, N., Babic, D., Nikolic, S., Plavsic, D., Amic, D. and Mihalic, Z. (1994). The Laplacian matrix in chemistry. *J. Chem. Inf. Comput. Sci.*, **34**, 368-376.

Wang, L. S.(1993). *Quantitative Structure-Activity Relationship of Organic Chemicals.* Beijing: China Environmental Science Press.

Wang L. S. (1997). *Molecualr Structure, Property, and Activity.* Beijing: Chemical Industry Press.

Wiener, H. (1947). Structural determination of paraffin boiling points. *J. Am. Chem. Soc.*, **69**, 17-20.

Xin H. W. (1991). *Molecular Topology.* He Fei: China Science and Technology University press.

Xu, L. and Hu, C. Y. (2001). *Applied Chemical Graph Theory*. Beijing: Science Press.

Yee, W. T., Sakamoto, K. and Ihaya, Y. J. (1977). Information theory of molecular properties. I. A theoretical study of the information content of organic molecules. *Rept. Univ. Electrocomm.*, **27**, 53-63.

Yu, X. Z, Lu, X., Ya, P. W. and Bai, L. L. (1999). A QSAR study of the antiallergic activities of substituted benzamides and their structures. *Chemometrics and Intelligent Laboratory Systems*, **45**, 95-100.

Yuan, H. and Parrill, A.L. (2000) QSAR development to describe HIV-1 integrase inhibition. *Journal of Molecular Structure (THEOCHEM)*, **529**, 273-282.

Zhao, Y. H., Ji, G. D., Cronin, M. T. D. and Dearden, U. J. C. (1998). QSAR study of the toxicity of benzoic acids to Vibrio fischeri, Daphnia magna and carp. *The Science of the Total Environment*, **216**, 205-215.

Zhou, J. J and Wang, T. (2001). *Methods of molecular modeling in drug design*. Beijing: Science Press.

Received May 20, 2003; accepted August 3, 2003

Qian-Nan Hu
Institute of Chemometrics and Intelligent Analytical Instruments
College of Chemistry and Chemical Engineering Central South University
Changsha, 410083, P.R. China
qnhu@263.net

Yi-Zeng Liang
Institute of Chemometrics and Intelligent Analytical Instruments
College of Chemistry and Chemical Engineering Central South University
Changsha, 410083, P.R. China
yzliang@public.cs.hn.cn

Kai-Tai Fang
Department of mathematics
Hong Kong Baptist University
Hong Kong, P. R. China
ktfang@hkbu.edu.hk