Developing Molecular Identification Numbers by an All-Paths Method

Chang-Yu Hu and Lu Xu*

Applied Spectroscopy Laboratory, Changchun Institute of Applied Chemistry, Academia Sinica, Changchun 130022, Jilin, Peoples Republic of China

Received June 28, 1996[⊗]

A new topological index is devised from an all-paths method. This molecular topological index has highly discriminating power for various kinds of organic compounds such as alkane trees, complex cyclic or polycyclic graphs, and structures containing heteroatoms and thus can be used as a Molecular IDentification number (MID) for chemical documentation. Some published MIDs derived from an all-paths method and their structural selectivity for alkane trees are also reviewed.

1. INTRODUCTION

Topological indices (graph invariants) are single numbers for the characterization of molecular structures. Two goals in developing topological indices are followed: (i) design of a useful descriptor for QSAR/QSPR[1−5] and (ii) design of a useful descriptor for chemical documentation, thus allowing storage and retrieval.[6−14] These two goals do not parallel one another. In QSAR/QSPR studies, the good correlations between the generated topological indices and the physical properties or biological activities were pursued. For the latter purpose, one wants to maximally discriminate among structures and expects to find a single mathematical invariant to represent a molecule uniquely. Many topological indices derived from the all-paths method had highly discriminatory power.[9−13] Those topological indices are Randic's molecular IDentification numbers, ID,[9] PID,[10] Balaban's BID,[11] Hall and Kier's τ,[12] and authors' ID.[13] In this article, we add a new one for documentation.

In a molecular graph, a path between two graph vertices i and j is any connected subgraph that starts and ends in those two vertices and traverses only once each of the intermediate vertices and edges. Chemical graphs are color graphs, so the path that starts at vertex i and ends at vertex j ($\text{Path}_{i,j}$) is not identical with the path that starts at vertex j and ends at vertex i($\text{Path}_{j,i}$) path. For example, in structure I, all paths that start at the eight atoms are listed Figure 1. A Path Identifier(PI) was devised to evaluate those paths, and the topological index was calculated from the PIs.

2. OUTLINE THE ALGORITHM

2.1. Vertex and Edge Properties. Chemical graphs are color graphs in which vertices are interpreted as distinct atoms, and the edges are colored by multiple connections.

In order to characterize the distinct atoms, the atom's connectivity δ (the number of non-hydrogen atoms attached to it) is the most commonly used vertex property for alkanes. The atomic number is introduced to distinguish the hetero-atoms. So we define a new atom invariant δ',

$$\delta' = \delta \times \sqrt{Z} \tag{1}$$

where Z is the atomic number.

The multiple bonds are single, double, triple bond, and aromatic bond which are coded as 1, 2, 3, and 1.5, respectively.

2.2. Path Identifier. The path can be weighted by a Path Identifier (PI). Before introducing our new path identifier, we review briefly some published schemes.

By assigning the weighting factors $\chi = (\text{m} \times \text{n})^{-1/2}$ to individual bonds, Randić[15] suggested the term w(pl) by means of

$$w(\text{pl}) = \Pi\chi = \prod^{1}(m \times n)^{-1/2} \tag{2}$$

Where m and n are the connectivity valences of the end points of the bond(m,n) in the path. We referred this w(pl) as PI01. The molecular IDentification number, ID, was generated by the sum of the w(pl) for all paths.[9] This ID numbers was also referred as Connectivity ID numbers, CID.[16]

Lately, Randić[10] used the prime numbers (2,3,5,6,11,13,-17,19,23) in replacement of the values of $m \times n$ in the original w(pl) formula, thus the path identifier PI02 and the prime ID numbers, PID, were devised.

Balaban[11] suggested a different weighted approach namely to replace the vertex degree m and n in the original w(pl) formula by distance sum d_i and d_j, and the path identifier PI03 and the molecular identification number BID were proposed.

Hall and Kier[12] introduced the term t_{ij} to weight the path between nodes i and j,

$$t_{ij} \rightarrow \text{GM}_{ij} \times \text{f}(n_{ij}) = \frac{(\prod_{k=1}^{n_{ij}} \delta_k^{\nu})^{1/n_{ij}}}{n_{ij}} \tag{3}$$

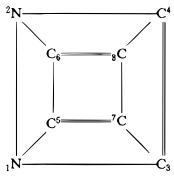
Where, GM stands for geometric mean, f(n_{ij}) is a function of the graph separation, k represents all the nodes along a path, δ^{ν} is the molecular connectivity valence, and n_{ij} is the number of nodes in the path. The topological index τ was calculated from the path identifier t_{ij}.

After carefully examining the path identifiers PI01, PI02, PI03, and t_{ij}, it can be seen that two important pieces of information are absent in those equations for the path of molecular graphs. They are (1) the position of the atoms along the path and (2) the multiple bonds and their position.

[⊗] Abstract published in *Advance ACS Abstracts,* November 1, 1996.

S0095-2338(96)00048-0 CCC: $14.00

So the following different paths in structure II have the identical PI01, PI02, PI03, and t_{ij} values.



II

1−5−6−2−4−8−7−3 (N−C−C−N−C−C−C−C)
1−5−6−2−4−3−7−8 (N−C−C−N−C=C−C−C)
1−2−4−3−7−8−6−5 (N−N−C=C−C−C=C−C)
1−2−4−8−6−5−7−3 (N−N−C−C=C−C=C−C)
1−2−6−5−7−8−4−3 (N−N−C−C=C−C−C=C)
1−2−6−8−4−3−7−5 (N−N−C=C−C=C−C=C)

In our early papers,[13,17] a term PI was defined as

$$\mathrm{PI} = \frac{\sum_{k=1}^{n_{ij}} N[k] b_{(k,k-1)} \times 10^{(-k)}}{n_{ij}^3} \tag{4}$$

where k is the sequence number of the nodes along the path between nodes i and j, n_{ij} is the total nodes in the path, $N[k]$ is the index number in node library, $b_{(k,k-1)}$ is the code for the bond between nodes k and $k - 1$. This is the path identifier PI05, and the molecular identification number MID05 can be derived from PI05.

Although the PI05 contains the $b_{(k,k-1)}$ factor to characterize the multiple bonds and the k factor to indicate the position of the atoms and bonds, the index number in node library $N[k]$ has no relation with the atom's properties, and some digital problems might be caused from the factor $10^{(-k)}$. So, PI05 is not a well-defined path identifier. Thus, in this study, a new path identifier is proposed, and here it is named as PI06

$$\mathrm{PI06} = \prod_{k_2}^{n_{ij}} \sqrt{\frac{b_{(k,k-1)}}{k} \times \frac{1}{\delta'[k] \times \delta'[k-1]}} \tag{5}$$

where k is the sequence number of the nodes along the path between nodes i and j, n_{ij} is the total nodes in the path, $b_{(k,k-1)}$ is the code for the bond between nodes k and $k - 1$, and $\delta'[k]$ is the revised atom's connectivity valence defined by eq 1.

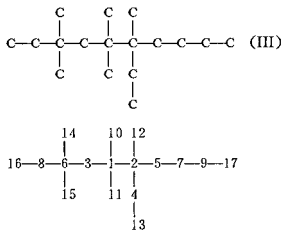
2.3. Atomic IDentification Number. For each atom, the atomic index is obtained by adding the PIs of all paths started at that vertex. Those atomic indices are the characteristic of atomic environments, e.g., the typically nonequivalent atoms in a molecule will in this way be assigned different numbers, and atoms in different molecules show different numbers. So, they are Atomic IDentification numbers (AIDs). The AID is defined as the sum of the PIs, which is by means of

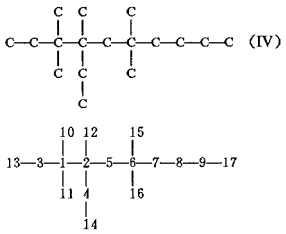
$$\mathrm{AID} = \sum \mathrm{PI} \tag{6}$$

AIDs calculated from the path identifiers t_{ij} and PI05 had been incorporated into the algorithm for computer perception of topological symmetry.[12,17]

2.4. Molecular IDentification Numbers. According to our tests, if the set $\{x_i\}$ $(i = 1, ..., N)$ is not equal to the set $\{y_i\}$ $(i = 1, ..., N)$, the opportunity of $\sum_{i=1}^{N} x_i^2 = \sum_{i=1}^{N} y_i^2$ is much less than that of $\sum_{i=1}^{N} x_i = \sum_{i=1}^{N} y_i$ in most cases.

For example, the set of AIDs of the 17 atoms in structure III, which is calculated from the sum of Randić's PI01, is not equal to that in another structure IV, but the molecular identification numbers CIDs of those two structures are equivalent. If $\mathrm{CID}' = \sum_{i=1}^{N} \mathrm{AID}[i]^2$, then the CID′ values





AID[1] = 3.3751905204503085	AID[1] = 3.3016181619379807
AID[2] = 3.4545212152358613	AID[2] = 3.4431474330271450
AID[3] = 2.8737432901339908	AID[3] = 2.6610167291357349
AID[4] = 2.7150761220612645	AID[4] = 2.7110548827975030
AID[5] = 2.9072872969075925	AID[5] = 2.9217960838455728
AID[6] = 3.2941309006906567	AID[6] = 3.3790770414117017
AID[7] = 2.8438086913987517	AID[7] = 2.9806137534515530
AID[8] = 2.6583695825354856	AID[8] = 2.9304719196707309
AID[9] = 2.7022344315892863	AID[9] = 2.6955660457252764
AID[10] = 2.4375952602251543	AID[10] = 2.4008090809689904
AID[11] = 2.4375952602251543	AID[11] = 2.4008090809689904
AID[12] = 2.4772606076179304	AID[12] = 2.4715737165135723
AID[13] = 2.4198487373471940	AID[13] = 2.3816229740227239
AID[14] = 2.3970654503453286	AID[14] = 2.4170052917950153
AID[15] = 2.3970654503453286	AID[15] = 2.4395385207058506
AID[16] = 2.3797511587108930	AID[16] = 2.4395385207058506
AID[17] = 2.4107682909325598	AID[17] = 2.4060530300685503
CID = 46.1813122667527480	CID = 46.1813122667527480

of structures III and IV are 127.558 626 385 037 46 and 127.559 687 807 099 73, respectively. They are not equal.

So, we define our new molecular identifications number as the sum of AID^2,

$$\mathrm{MID} = \sum_{i=1}^{N} \mathrm{AID}[i]^2 \tag{7}$$

The calculation of our newly proposed molecular identification number MID06 is illustrated in Figure 1.

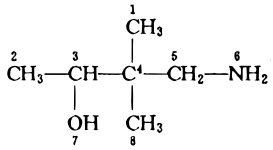
3. TESTING OF THE STRUCTURAL SELECTIVITY

Since we have built an exhaustive and irredundant structural isomer generator,[18,19] we can produce all members of a given family of chemical structures. Thus the discriminatory power of the proposed topological indices can be tested in a more systematic manner. The testing is carried out in the following families of structures: (i) alkane trees and (ii) structures containing heteroatoms.

3.1. Alkane Trees. The acyclic alkanes are trees with no vertex having a degree higher than 4. In this paper, only constitutional alkanes are included, and the steric hindrance is not considered.

To test the uniqueness of the MID numbers for alkane trees, the molecular formula C_nH_{2n+2} $(n = 1, ..., 22)$ was supplied to our isomer generator to produce all members of the alkane tree family with up to 22 atoms, and a total of 380 743 4 isomers were obtained.

We have examined all the molecular identification numbers (derived from the all-paths method) of those 3 807 434 isomers; the testing results are listed in Table 1. In Table



(I)

Atom 1	PI06
1;1 CH_3	1.
1;2 $CH_3-C-CH-CH_3$	0. 00231481481481
1;3 CH_3-C-CH	0. 01701034543599
1;4 CH_3-C	0. 20412414523193
1;5 CH_3-C-CH_2	0. 02083333333333
1;6 $CH_3-C-CH_2-NH_2$	0. 00334095645207
1;7 $CH_3-C-CH-OH$	0. 00215417791459
1;8 CH_3-C-CH_3	0. 02946278254944
AID[1]=	1. 27924055573217

Atom 2	PI06
2;1 $CH_3-CH-C-CH_3$	0. 00231481481481
2;2 CH_3	1.
2;3 CH_3-CH	0. 23570226039552
2;4 CH_3-CH-C	0. 01964185503296
2;5 $CH_3-CH-C-CH_2$	0. 00163682125275
2;6 $CH_3-CH-C-CH_2-NH_2$	0. 00022732329335
2;7 $CH_3-CH-OH$	0. 03655761147090
2;8 $CH_3-CH-C-CH_3$	0. 00231481481481
AID[2]=	1. 29839550107510

Atom 3	PI06
3;1 $CH-C-CH_3$	0. 01701034543599
3;2 $CH-CH_3$	0. 23570226039552
3;3 CH	1.
3;4 $CH-C$	0. 11785113019776
3;5 $CH-C-CH_2$	0. 01202813060812
3;6 $CH-C-CH_2-NH_2$	0. 00192890210695
3;7 $CH-OH$	0. 01701034543599
3;8 $CH-C-CH_3$	0. 21934566882542
AID[3]=	1. 62087678300575

Atom 4	PI06
4;1 $C-CH_3$	0. 20412414523193
4;2 $C-CH-CH_3$	0. 01964185503296
4;3 $C-CH$	0. 11785113019776
4;4 C	1.
4;5 $C-CH_2$	0. 14433756729741
4;6 $C-CH_2-NH_2$	0. 02834895555488
4;7 $C-CH_2-OH$	0. 01827880573545
4;8 $C-CH_3$	0. 20412414523193
AID[4]=	1. 73670660428232

Atom 5	PI06
5;1 CH_2-C-CH_3	0. 02083333333333
5;2 $CH_2-C-CH-CH_3$	0. 00163682125275
5;3 CH_2-C-CH	0. 01202813060812
5;4 CH_2-C	0. 14433756729741
5;5 CH_2	1.
5;6 CH_2-NH_2	0. 27776190340118
5;7 $CH_2-C-CH-OH$	0. 00152323381129
5;8 CH_2-C-CH_3	0. 02083333333333
AID[5]=	1. 47895432303740

Atom 6	PI06
6;1 $H_2N-CH_2-C-CH_3$	0. 00334095645207
6;2 $N2N-CH_2-C-CH-CH_3$	0. 00022732329335
6;3 H_2N-CH_2-C-CH	0. 00192890210695
6;4 H_2N-CH_2-C	0. 02834895555488
6;5 H_2N-CH_2	0. 27776190340118
6;6 NH_2	1.
6;7 $H_2N-CH_2-C-CH-OH$	0. 00021154816137
6;8 $H_2N-CH_2-C-CH_3$	0. 00334095645207
AID[6]=	1. 31516054542187

Atom 7	PI06
7;1 $OH-CH-C-CH_3$	0. 00215417791459
7;2 $OH-CH-CH_3$	0. 03655761147090
7;3 $OH-CH$	0. 21934566882542
7;4 $OH-CH-C$	0. 01827880573545
7;5 $OH-CH-C-CH_2$	0. 00152323381129
7;6 $OH-CH-C-CH_2-NH_2$	0. 00021154816137
7;7 OH	1.
7;8 $OH-CH-C-CH_3$	0. 00215417791459
AID[7]=	1. 28022522383361

Atom 8	PI06
8;1 CH_3-C-CH_3	0. 02946278254944
8;2 $CH_3-C-CH-CH_3$	0. 00231481481481
8;3 CH_3-C-CH	0. 01701034543599
8;4 CH_3-C	0. 20412414523193
8;5 CH_3-C-CH_2	0. 02083333333333
8;6 $CH_3-C-CH_2-NH_2$	0. 00334095645207
8;7 $CH_3-C-CH-OH$	0. 00215417791459
8;8 CH_3	1.
AID[8]=	1. 27924055573217

$$\mathrm{MID06} = \sum_{i=1}^{8} \mathrm{AID}[i]^2 = 16.15806482472203$$

Figure 1. The calculation of MID06 of structure I.

1, the molecular identification numbers CID, PID, BID, τ, MID05, and MID06 have been described above, and the CID′, PID′, BID′, τ', and MID05′ are calculated from eq 6 and 7 with the path identifiers PI01, PI02, PI03, t_{ij}, and PI05. From Table 1, it can be seen that all the molecular identification numbers derived from the all-paths method have highly discriminating power for alkanes. Our newly proposed molecular identification number MID06 has no degeneracy for alkane trees up to 22 carbon atoms.

3.2. Structures Containing Heteroatoms. Since more than 80% of the known chemical structures contains heteroatoms, it is a much more important task to process structural information on heteroatoms for graph theoretical approaches, and that represents a more challenging group, especially to those with cyclic or polycyclic systems. In order to test the uniqueness of the MID06 numbers for those

Table 1. Distribution of Duplicate MIDs Numbers Derived from an All-Path Method for Alkane Trees[a]

n	no. of alkane isomers	CID	CID′	PID	PID′	BID	BID′	τ	τ'	MID05	MID05′	MID06
1	1	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	0	0	0	0	0	0	0	0	0
3	1	0	0	0	0	0	0	0	0	0	0	0
4	2	0	0	0	0	0	0	0	0	0	0	0
5	3	0	0	0	0	0	0	0	0	0	0	0
6	5	0	0	0	0	0	0	0	0	0	0	0
7	9	0	0	0	0	0	0	0	0	0	0	0
8	18	0	0	0	0	0	0	0	0	0	0	0
9	35	0	0	0	0	0	0	0	0	0	0	0
10	75	0	0	0	0	0	0	0	0	0	0	0
11	159	0	0	0	0	0	0	0	0	0	0	0
12	355	0	0	0	0	0	0	0	0	0	0	0
13	802	0	0	0	0	0	0	0	0	0	0	0
14	1859	0	0	0	0	0	0	0	0	0	0	0
15	4347	1	0	0	0	0	0	0	0	0	0	0
16	10359	1	0	0	0	0	0	0	0	0	0	0
17	24894	3	0	0	0	0	0	0	0	0	0	0
18	60523	10	0	0	0	0	0	0	0	1	1	0
19	148284	23	0	0	0	0	0	1	0	1	1	0
20	366319	88	0	1	0	1	1	3	0	6	1	0
21	910726	227	0	3	0	2	0	5	0	7	6	0
22	2278658	666	0	12	1	62	14	6	0	57	48	0

[a] Molecular formula: C_nH_{2n+2}.

structures, an exhaustive set of a total of 430 472 structures having $N = 8$ atoms with up to one oxygen or/and one nitrogen heteroatoms were selected. Those structures are exhaustively generated from the following molecular formulas:

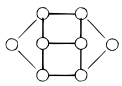
(1) C_8H_n $n = 18,16,14,12,10,8,6,4,2,0$

(2) C_7H_nO $n = 16,14,10,8,6,4,2,0$

(3) C_7H_nN $n = 17,15,13,11,9,7,5,3,1$

(4) C_6H_nNO $n = 15,13,11,9,7,5,3,1$

The molecular formulas and their isomer numbers are listed in Table 2. Those structures include various types of structures such as saturated and unsaturated structures or acyclic, cyclic, and polycyclic structures. Many of them have the same graph core with only different locations of heteroatoms or/and multiple bonds. For example, there are 261 structures with the same graph core



(Those 261 structures are listed in ref 14).

After calculating the MID06 numbers of those 430 472 structures, no duplicate with an identical MID06 number was found. Thus the MID06 numbers are highly discriminatory topological indices for structures containing heteroatoms.

All the algorithms are programmed in ANSI C language and run on an IBM PC computer and SGI workstations. All calculations are carried out with ANSI C double precision.

Table 2. Some Molecular Formula (Having $n = 8$ Vertexes) and Their Isomer Numbers

molecular formula	no. of isomers	total no. of isomers
C_8H_{18}	18	18
$C_7H_{16}O$	72	90
$C_7H_{17}N$	89	179
$C_6H_{15}NO$	405	584
C_8H_{16}	139	723
$C_7H_{14}O$	596	1319
$C_7H_{15}N$	801	2120
$C_6H_{13}NO$	3418	5538
C_8H_{14}	654	6192
$C_7H_{12}O$	2589	8781
$C_7H_{13}N$	3826	12607
$C_6H_{11}NO$	14410	27017
C_8H_{12}	2082	29099
$C_7H_{10}O$	7166	36265
$C_7H_{11}N$	11773	48038
C_6H_9NO	37202	85240
C_8H_{10}	4679	89919
C_7H_8O	13177	103096
C_7H_9N	24627	127723
C_6H_7NO	61255	188978
C_8H_8	7437	196415
C_7H_6O	15804	212219
C_7H_7N	34745	246964
C_6H_5NO	61974	308938
C_8H_6	7982	316920
C_7H_4O	11332	328252
C_7H_5N	31163	359415
C_6H_3NO	33896	393311
C_8H_4	5308	398619
C_7H_2O	3971	402590
C_7H_3N	15489	418097
C_6H_1NO	7038	425117
C_8H_2	1804	426921
C_7O	356	427277
C_7HN	2991	430268
C_8	204	430472

4. DISCUSSIONS

These molecular ID numbers generated from an all-paths algorithm highly condense some important structural features into a single number that parallels some molecular properties. Though it is difficult to draw some distinct physicamchemical meaning from those numbers, we can see that different molecules have distinct ID values. They can represent the molecule into a single number uniquely. This is of great convenience in chemical documentation, structure storage, and retrieval.

To generate highly discriminative ID numbers from all-paths algorithm, the following factors should be of concern:

(1) Some proper atom properties (or vertex properties) should be selected to characterize the atoms. The atom's connectivity δ, the molecular connectivity valence δ^v introduced by Hall and Kier, the distance sum d_i, the covalent radiis, and the atomic number Z are all useful properties. For chemical structures, e.g., color graphs, the selected atom property should be able to distinguish the different atoms and their covalent valences, such as CH_3-, $-CH_2-$, >CH−, −CH═, >C<, >C═, ═C═, >N−, −N═, etc. The delicate atom property can be helpful for the discriminatory power, for example, PID and BID are greater in discrimination among alkanes than CID. From section 2.2, it can be generated that all PI equations contain a factor of $\Pi_{k=1}^{n_{ij}} p_k$, if the atom property p_k value is too low or too high, the PI might lose some atoms' information for a long path owing to the numerical problem. So we combine the δ and Z together as $\delta' = \delta \times \sqrt{Z}$, not $\delta \times Z$ or other forms, in case that some atom's δ' is too high or too low.

(2) The multiple bonds, heteroatoms, and their positions' information along the path should be incorporated to the path identifier PI. So our new PI06 contains the factor $b_{(k,k-1)}/k$. Thus it can distinguish the paths of structure II listed in section 2.2. So our new topological index MID06 can discriminate among the structures in Table 2, but the other MIDs numbers such as CID, PID, BID, and τ had low discriminatory power for these structures.

ACKNOWLEDGMENT

The authors acknowledge the financial support of the National Natural Science Foundation of China and partial financial support of the Yanbian Aodong Pharmaceutical Groups of Jilin province.

REFERENCES AND NOTES

(1) Bonchev, D. Toplogical Order in Molecules 1. Molecular Branching Revisited. *J. Mol. Struct.* (THEOCHEM) **1995**, *336*, 137−156.

(2) Randić, M. Novel Graph Theoretical Approach to Heteroatoms in Quantitative Structure−Activity Relationships. *Chemom. Intell. Lab. System*. **1991**, *10*, 213−227.

(3) Xu, L.; Wang, H. Y.; Su, Q. A Newly Proposed Topological Index for Discrimination of Cis/Trans Isomers and for the Studies of QSAR/QSPR. *J. Comput. Chem*. **1992**, *16*, 187−194.

(4) Yang, Y. Q.; Xu, L.; Hu, C. Y. Extended Adjacency Matrix Indices and Their Applications. *J. Chem. Inf. Comput. Sci*. **1994**, *34*(5), 1140−1145.

(5) Xu, L.; Yao, Y. Y.; Wang, H. M. New Topological Indices and Prediction of Phase Transfer Energy for Protorated Amines and Tetraalkylamines Ions. *J. Chem. Inf. Comput. Sci*. **1995**, *35*(1), 45−49.

(6) Burden, F. R. Molecular IDentification Number for Substructure Searches. *J. Chem. Inf. Comput. Sci*. **1989**, *29*, 225−227.

(7) Elk, S. B.; Gutman, I. Further properties derivable from the Matula numbers of an alkane. *J. Chem. Inf. Comput. Sci*. **1994**, *44*, 331−334.

(8) Muller, W. R.; Szymanski, K.; Knop, J. V.; Mihalic, Z.; Trinajstić, N. The Walk ID Numbers Revisited. *J. Chem. Inf. Comput. Sci*. **1993**, *33*, 231−233.

(9) Randić, M. On Molecular IDentification Numbers. *J. Chem. Inf. Comput. Sci*. **1984**, *24*, 164−175.

(10) Randić, M. Molecular ID Numbers: by Design. *J. Chem. Inf. Comput. Sci*. **1986**, *26*, 134−136.

(11) Balaban, A. T. Numerical Modeling of Chemical Structures: Local Graph Invariants and Topological Indices. In *Graph Theory and Topology*; King, R. B., Rouvray, D. H., Eds.; Elsevier: Amsterdam, 1987; pp 159−176.

(12) Hall, L. H.; Kier, L. B. Determination of Topological Equivalence in Molecular Graphs from the Topological State. *Quant. Struct.-Act. Relat*. **1990**, *9*, 115−131.

(13) Hu, C. Y.; Xu, L. A New Topological Index for CIAC-^{13}CNMR Information System. *Anal. Chem. Acta* **1995**, *318*, 117−123.

(14) Hu, C. Y.; Xu, L. On Highly Discriminating Topological Index. *J. Chem. Inf. Comput. Sci*. **1996**, *36*, 82−90.

(15) Randić, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc*. **1975**, *97*, 6609−6615.

(16) Randić, M.; Trinajstić, N. *J. Mol. Struct*. **1993**, *300*, 551.

(17) Hu, C. Y.; Xu, L. Study on Computer Perception of Topological Symmetry. *Acta Chem. Sinica (Chinese)*, in press.

(18) Hu, C. Y.; Xu, L. Expert System for Elucidation of Structures of Organic Compounds. *Sci. China, Ser. B* **1995**, *38*(3), 296−304.

(19) Hu, C. Y.; Xu, L. Principles for Structure Generation of Organic Isomers from Molecular Formula. *Anal. Chim. Acta* **1994**, *298*, 75−85.

CI960048P