On Hall and Kier's Topological State and Total Topological Index

Chang-Yu Hu† and Lu Xu*

Applied Spectroscopy Laboratory, Changchun Institute of Applied Chemistry, Academia Sinica, Changchun 130022, Jilin, People's Republic of China

Received May 18, 1994⊗

The topological state and the total topological index, τ, have been tested systematically. Counterexamples are provided proving that the topological state method cannot determine classes of symmetrically equivalent atoms in a molecule fully correct. The search for counterexamples to the total topological index τ in the field of alkane trees up to 20 atoms produces five duplications. Thus τ is a highly discriminating index but not unique.

Hall and Kier have recently introduced a new vertex invariant, named the topological state (S), and a new topological index, named the total topological index (τ), in the journal of *Quant. Struct.-Act. Relat.*[1] The S and τ were defined in the following ways: The topological state matrix **T** was proposed which contains entries t_{ij}, the numerical information which characterizes the graph relation between nodes i and j

$$t_{ij} \rightarrow [\mathrm{GM}_{ij}]^a[\mathrm{f}(n_{ij})]^b = [(\prod_{k=1}^{n_{ij}} \delta_k^{\mathrm{v}})^{1/n_{ij}}]^a[n_{ij}]^b$$

$$\delta^{\mathrm{v}} = Z^{\mathrm{v}} - h$$

$$n_{ij} = d_{ij} + 1$$

Where, GM stands for geometric mean, $\mathrm{f}(n_{ij})$ is a function of the graph separation, k represents all the nodes along a path, a is either ± 1, and b is 0, ± 1, ± 2, or ± 3; for each skeletal atom represented by the node, δ^{v} is the molecular connectivity valence, Z^{v} is the number of valence electrons, h is the number of hydrogen atoms, and d_{ij} is the distance between nodes i and j.

Then S_i^{v} is obtained as the row sum of entries in matrix **T**

$$S_i^{\mathrm{v}} = \sum_{j=1}^{N} t_{ij}$$

For a tree graph each matrix element consists of a single entry. For cyclic graphs there can be more than one entry, corresponding to the multiple paths. Hence, for cyclic graphs each topological state matrix element, t_{ij}, is a sum of terms for all paths between nodes i and j. It should be noted that Hall and Kier used the symbol S_i for their new electro-topological state index in their later paper[2] and referred to the symbol T_i in place of S_i for the topological state.[3] But, in this article, we still use S_i for the topological state corresponding to ref 1.

The total topological index τ is developed by summing the atomic topological state values, including the diagonal elements and the triangular half of the topological matrix.

$$\tau = \sum_{i=1}^{N} \sum_{j>=i}^{N} t_{ij}$$

In Hall and Kier's studies, a family of algorithms have been investigated for which a is either ± 1 and b is 0, ± 1, ± 2, or ± 3. However, in this paper, only the algorithm, $a = 1$, $b = -2$, is tested. It is programmed in Borland C++ and runs on an IBM PC 486/33 computer. The computations were carried out in long double precision (80 bytes).

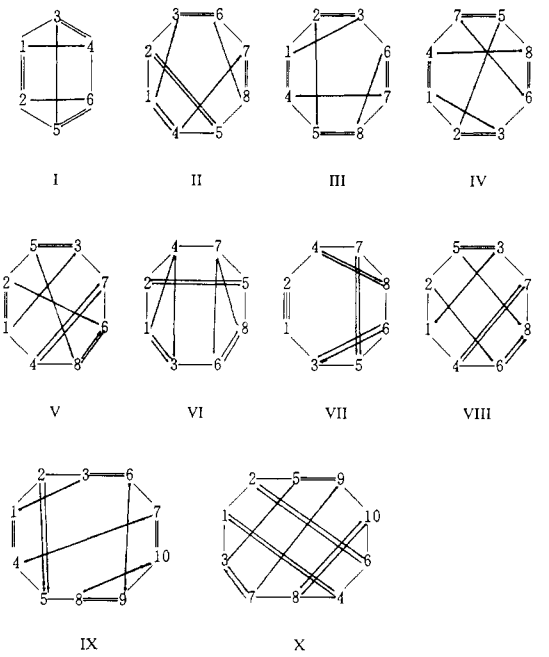


Figure 1. Counterexamples for topological state method.

TEST OF TOPOLOGICAL STATE METHOD

In Hall and Kier's studies, the topological equivalence in molecular graphs was determined by topological state, which nodes with identical topological state values are topological equivalent. A great amount of variety graphs are applied to test the method, and the topological equivalence in those molecular graphs was determined correctly by the topological state method. But, according to our tests, it is only a good approximation and not fully correct, because of their exit counterexamples for which the number of distinct symmetry types of atoms are greater than that computed by the topological state method.

Without going into the method details, it is not difficult to show that symmetrically distinct nodes in some graphs can give the same topological state value. From the form of t_{ij}, it can be seen that the δ^{v} values are used to weight nodes in the graphs, and the equation can characterize the paths with heteroatoms but fail to characterize some paths

† Program of ESESOC can be obtained from author C. Y. Hu.
⊗ Abstract published in *Advance ACS Abstracts,* August 15, 1994.

0095-2338/94/1634-1251$04.50/0

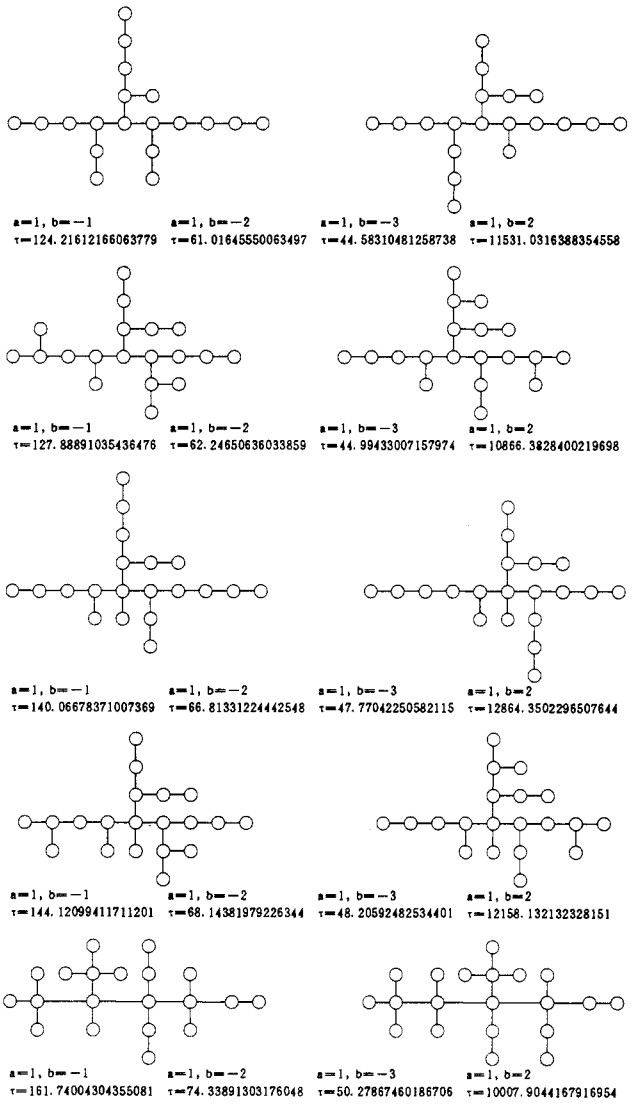
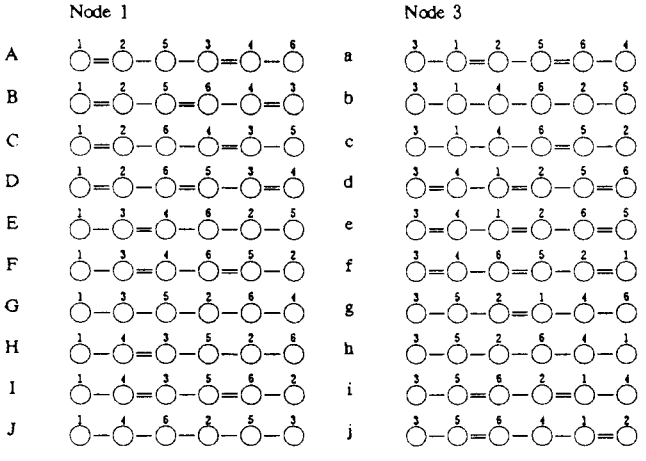


Figure 2. Five pairs of alkane trees with identical τ values.

with multiple bonds for its absence of the bond multiplicity information. Taking structure I of Figure 1 as an example, consider atoms 1 and 3, which are topologically nonequivalent. There are 10 paths in the set of graph node sequences with $n_{ij} = 6$, which terminated on node 1, and 10 paths in the set of graph node sequences with $n_{ij} = 6$, which terminated on node 3. The two sets are given below.



There are only six correspondences (B,*d*), (D,*e*), (F,*a*), (G,*b*),

Table 1. Computed Results of the Counterexamples by Topological State Method[a]

I	topological state values: $S_1^v = 15.517777777778$; $S_2^v = 15.517777777778$; $S_3^v = 15.517777777778$; $S_4^v = 15.517777777778$; $S_5^v = 15.517777777778$; $S_6^v = 15.517777777778$ sets of equivalent nodes: (a) (1,2,3,4,5,6); (b) (1,2), (3,4,5,6)
II	topological state values: $S_1^v = 21.221269841270$; $S_2^v = 21.221269841270$; $S_3^v = 20.933004535147$; $S_4^v = 20.831825396825$; $S_5^v = 20.831825396825$; $S_6^v = 20.933004535147$; $S_7^v = 21.221269841270$; $S_8^v = 21.221269841270$ sets of equivalent nodes: (a) (1,2,7,8), (3,6), (4,5); (b) (1,2), (3), (4,5), (6), (7,8)
III	topological state values: $S_1^v = 21.221269841270$; $S_2^v = 21.221269841270$; $S_3^v = 20.933004535147$; $S_4^v = 20.831825396825$; $S_5^v = 20.831825396825$; $S_6^v = 20.933004535147$; $S_7^v = 21.221269841270$; $S_8^v = 21.221269841270$ sets of equivalent nodes: (a) (1,2,7,8), (3,6), (4,5); (b) (1,8), (2,7), (3,6), (4,5)
IV	topological state values: $S_1^v = 21.686961451247$; $S_2^v = 21.686961451247$; $S_3^v = 21.686961451247$; $S_4^v = 21.499047619048$; $S_5^v = 21.499047619048$; $S_6^v = 21.499047619048$; $S_7^v = 21.787312925170$; $S_8^v = 21.787312925170$ sets of equivalent nodes: (a) (1,2,3), (4,5,6), (7,8); (b) (1), (2,3), (4), (5,6), (7,8)
V	topological state values: $S_1^v = 22.166269841270$; $S_2^v = 22.166269841270$; $S_3^v = 22.166269841270$; $S_4^v = 22.166269841270$; $S_5^v = 22.166269841270$; $S_6^v = 22.166269841270$; $S_7^v = 22.166269841270$; $S_8^v = 22.166269841270$ sets of equivalent nodes: (a) (1,2,3,4,5,6,7,8); (b) (1,2,3,5), (4,6,7,8)
VI	topological state values: $S_1^v = 27.561349206349$; $S_2^v = 26.615793650794$; $S_3^v = 26.615793650794$; $S_4^v = 26.183843537415$; $S_5^v = 26.615793650794$; $S_6^v = 26.615793650794$; $S_7^v = 26.183843537415$; $S_8^v = 27.561349206349$ sets of equivalent nodes: (a) (1,8), (2,3,5,6), (4,7); (b) (1,8), (2,5), (3,6), (4,7)
VII	topological state values: $S_1^v = 14.565986394558$; $S_2^v = 14.565986394558$; $S_3^v = 16.181326530612$; $S_4^v = 16.181326530612$; $S_5^v = 16.633724489796$; $S_6^v = 16.633724489796$; $S_7^v = 16.633724489796$; $S_8^v = 16.633724489796$ sets of equivalent nodes: (a) (1,2), (3,4), (5,6,7,8); (b) (1,2), (3,4), (5,7), (6,8)
VIII	topological state values: $S_1^v = 19.116028809352$; $S_2^v = 19.116028809352$; $S_3^v = 20.897013690384$; $S_4^v = 20.897013690384$; $S_5^v = 20.897013690384$; $S_6^v = 20.897013690384$; $S_7^v = 21.062404430426$; $S_8^v = 21.062404430426$ sets of equivalent nodes: (a) (1,2), (3,4,5,6), (7,8); (b) (1,2), (3,5), (4,6), (7,8)
IX	topological state values: $S_1^v = 29.449725371630$; $S_2^v = 29.324725371630$; $S_3^v = 28.766219450743$; $S_4^v = 28.376732174351$; $S_5^v = 27.913145628622$; $S_6^v = 27.913145628622$; $S_7^v = 28.376732174351$; $S_8^v = 28.766219450743$; $S_9^v = 29.324725371630$; $S_{10}^v = 29.449725371630$ sets of equivalent nodes: (a) (1,10), (2,9), (3,8), (4,7), (5,6); (b) (1), (2), (3), (4), (5), (6), (7), (8), (9), (10)
X	topological state values: $S_1^v = 30.799712773998$; $S_2^v = 30.799712773998$; $S_3^v = 30.799712773998$; $S_4^v = 30.799712773998$; $S_5^v = 30.799712773998$; $S_6^v = 30.799712773998$; $S_7^v = 30.799712773998$; $S_8^v = 30.799712773998$; $S_9^v = 30.799712773998$; $S_{10}^v = 30.799712773998$ sets of equivalent nodes: (a) (1,2,3,4,5,6,7,8,9,10); (b) (1,2,3,5), (4,6,7,9), (8,10)

[a] (a) Sets of equivalent nodes determined by topological state method and (b) the distinct sets of equivalent nodes.

Chart 1

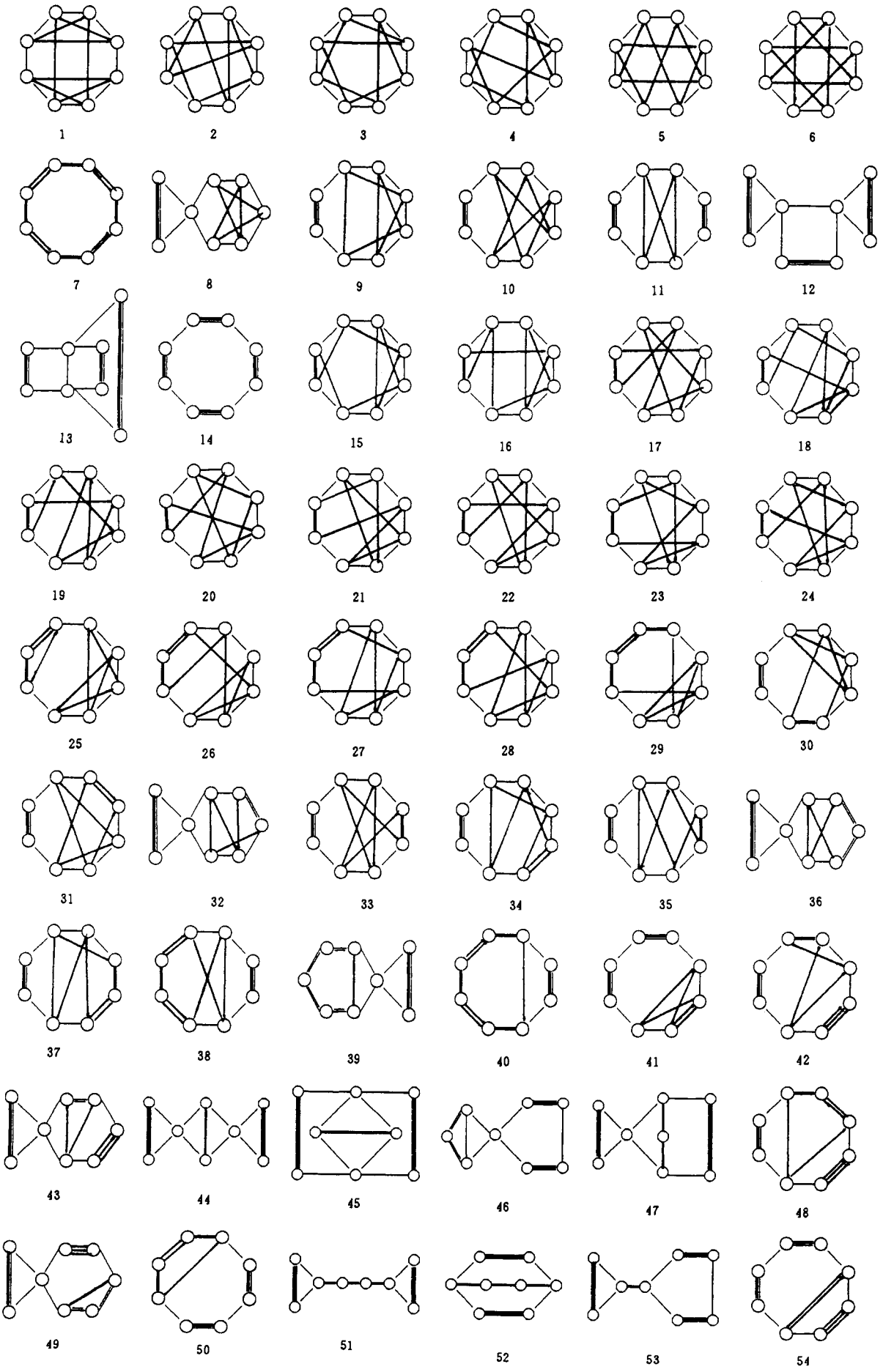

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54

Chart 2

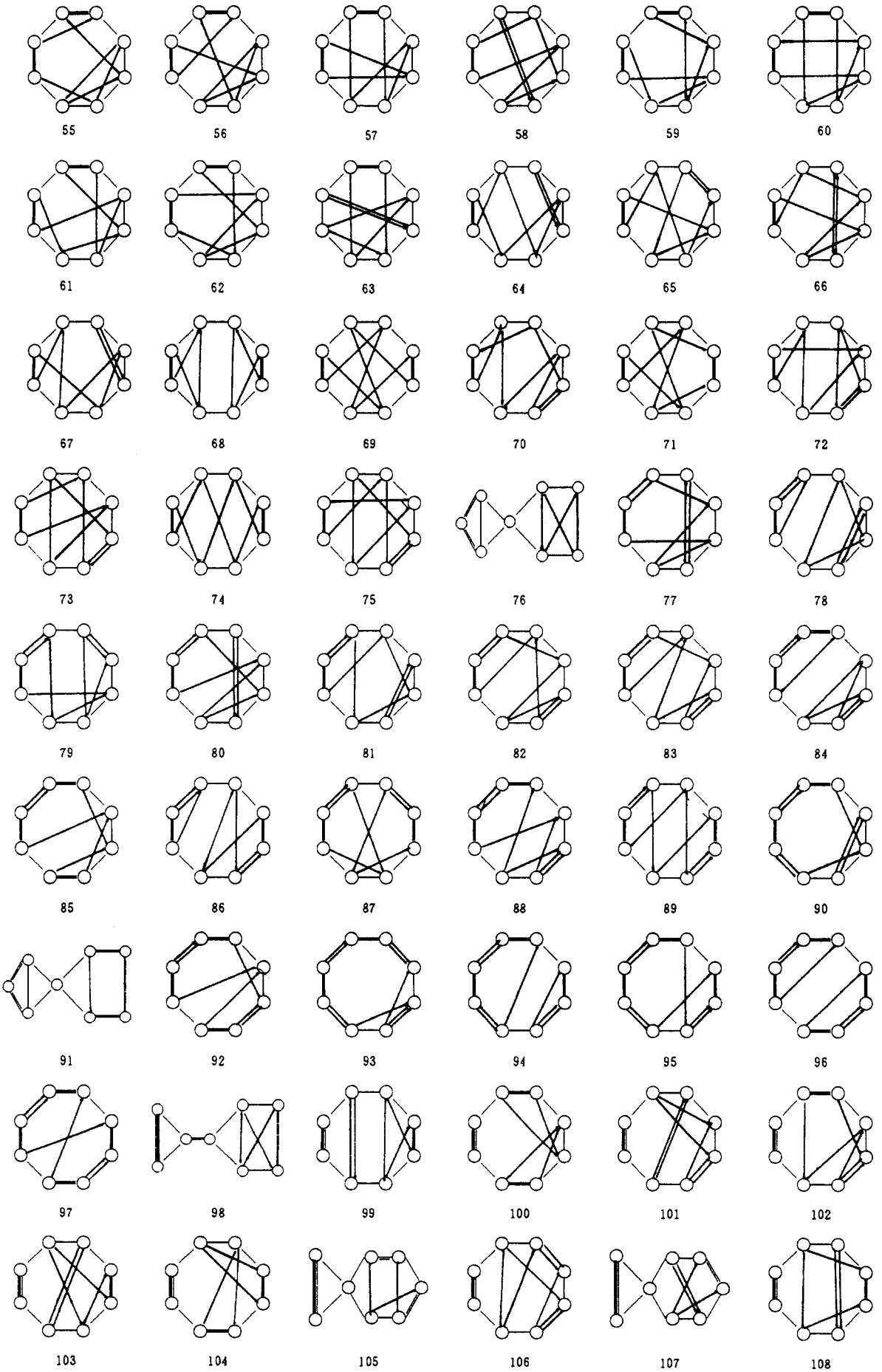

55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108

Chart 3

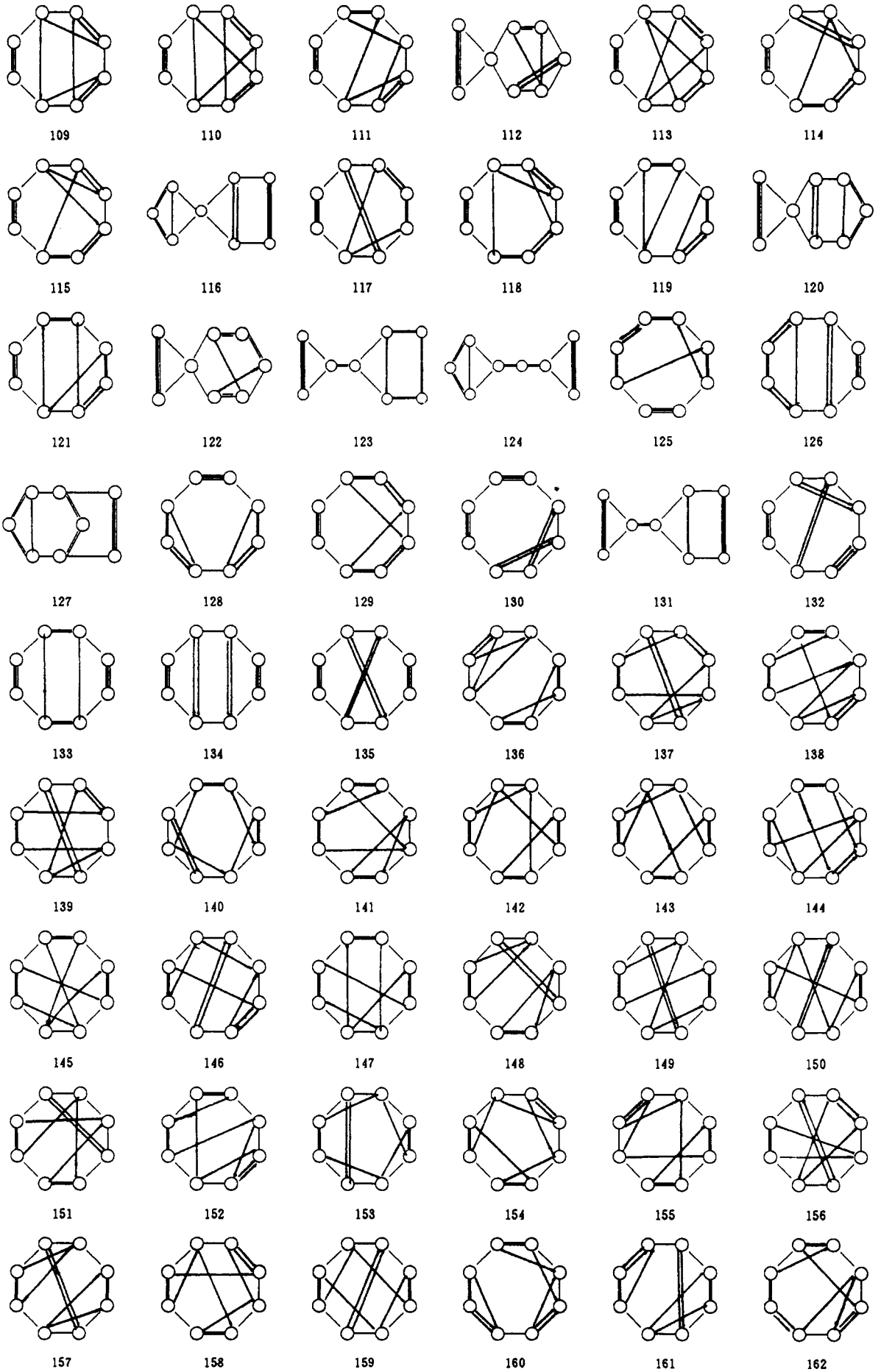
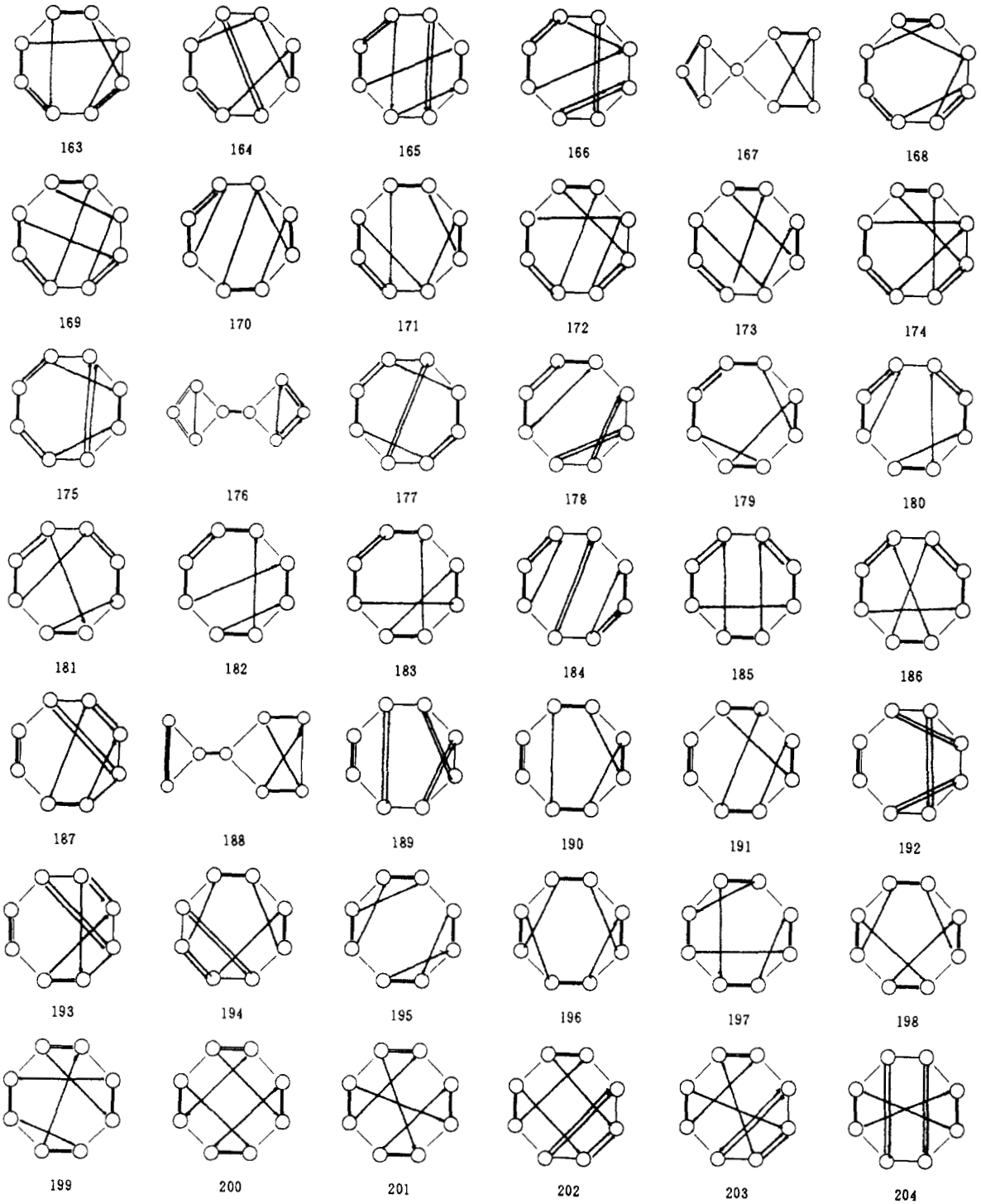

109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162

Chart 4



(I,*i*), and (J,*h*), so atoms 1 and 3 are topologically nonequivalent. But all the t_{ij} values of these 20 paths is 0.111 111 11 calculated by Hall and Kier's equation, thus Hall and Kier's method fails to determine the nonequivalence of nodes 1 and 3 structure I in Figure 1. Figure 1 shows several graphs which are counterexamples. The computed results of C. Y. Hu and L. Xu topological equivalent classes of atoms in these graphs are also listed. The distinct topological equivalent classes are calculated by our algorithms described in refs 4 and 5.

TEST OF TOTAL TOPOLOGICAL INDEX T

The total topological index τ is highly attractive for its high discrimination of structural isomers. It is unique for the graphs investigated by Hall and Kier. Is it unique for all other graphs? In this paper, we try to answer this question.

In our laboratory, an interactive program, ESESOC (expert system for the elucidation of the structures of organic compounds), has been developed to reduce comprehensive spectroscopic information to its structural implications.[6] The structure generator of ESESOC accepts specific type of information (molecular formula, substructure constraints, etc.) and produces an exhaustive and irredundant list of candidate structures.[7] The exhaustiveness and irredundancy of the structure generator of ESESOC has been shown.[7] So we have ready a set of programs for exhaustive and irredundant generation molecular graphs from a given segment set or a

Chart 5

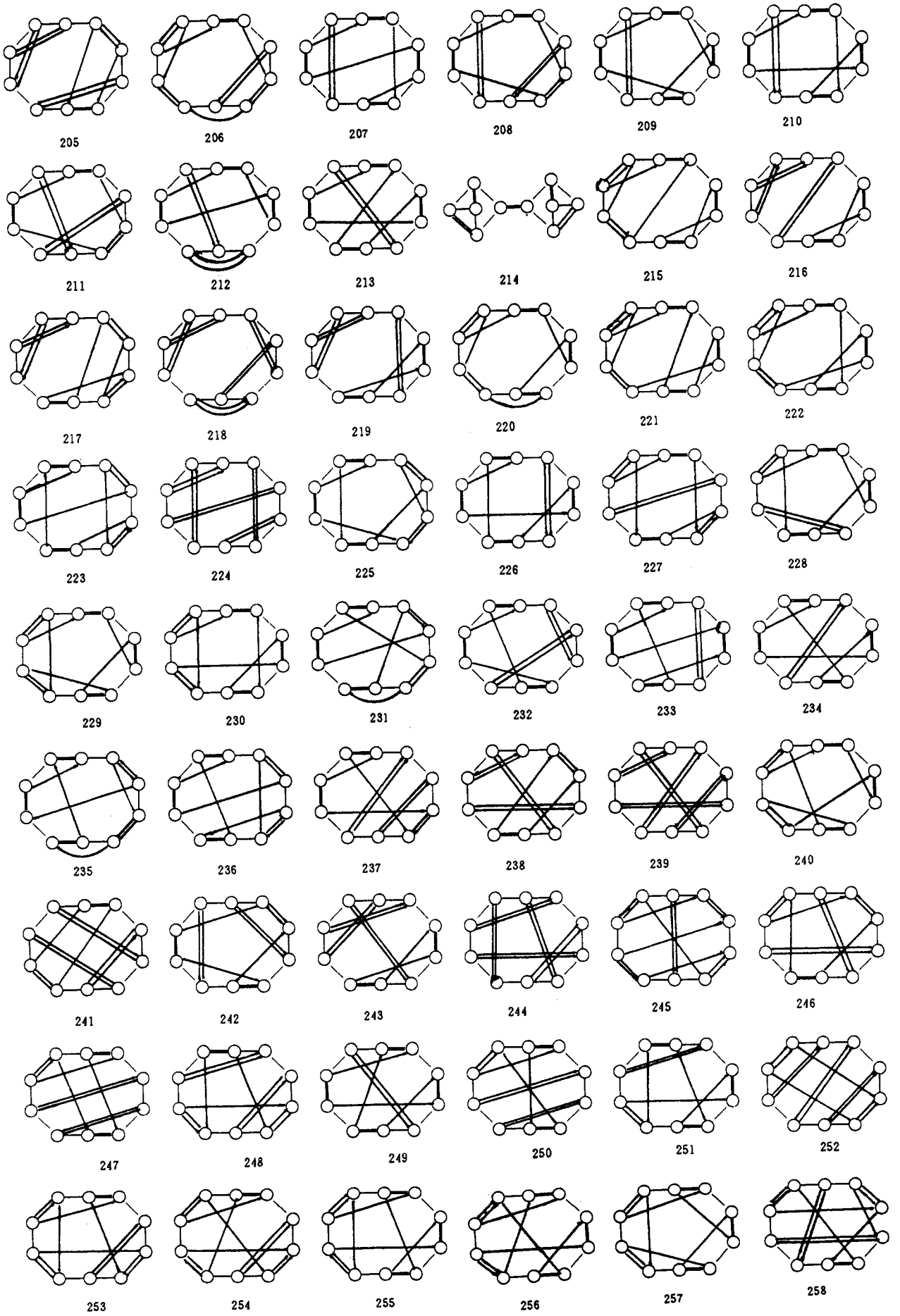

205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258

Table 2. Duplication Sets of Structures 1−204 Generated from C_8

duplication sets	τ
(7,14)	48.877 505 668 934
(40,50,54)	56.386 814 058 957
(84,99)	92.411 224 489 796
(91,116)	74.630 022 675 737
(93,130)	64.900 090 702 948
(96,126,133,134)	66.154 319 727 891
(97,135)	67.691 405 895 692
(108,109)	99.135 328 798 186
(123,131)	61.819 455 782 313
(125,132)	66.770 226 757 370
(143,144)	126.875 532 879 819
(149,150)	128.498 242 630 385
(153,154)	124.818 866 213 152
(165,173)	101.745 725 623 583
(175,179,192)	80.014 761 904 762
(178,189,190)	77.752 879 818 594
(182,187,191)	81.097 641 723 356
(183,193)	82.348 367 346 939
(194,197,198)	100.207 369 614 512
(195,196)	93.309 795 918 367
(200,202)	103.856 734 693 878
(201,203,204)	104.665 079 365 079

Table 3. τ Values of Structures 205−258

graphs	τ
(205,217,221)	151.496 942 554 800
(206,218,220)	150.618 890 148 652
(207,223,224,227)	163.830 547 996 977
(208,209,225,228,229)	163.696 090 954 900
(210,226,230)	168.223 858 654 573
(211,232,240)	168.049 800 957 420
(212,235)	163.414 884 731 670
(213,234,237,238,239)	171.906 920 508 944
(214)	121.762 166 792 643
(215,216)	142.012 892 416 226
(219,222)	155.023 469 387 755
(231)	168.123 023 431 595
(233,236)	170.324 160 997 732
(241,247,252)	173.998 563 869 992
(242,245,250,257)	174.573 041 068 279
(243)	172.858 911 564 626
(244,246,248,251,253,255)	176.840 070 546 737
(249,254,256)	178.138 354 749 307
(258)	180.489 984 882 842

molecular formula and programs for construction of various topological indexes[8−11]. We have computed the τ values of various graphs.

First, all alkane trees with up to 20 atoms (a total of 618 050 alkanes) are exhaustively and irredundantly generated from molecular formula C_nH_{2n+2} (n = 1−20) by ESESOC, and the τ values of these alkane trees are computed and examined. Five pairs of alkanes with the identical τ value are found (see Figure 2), two pairs are in the nondecane family, and three pairs are in the eicosane family. The other algorithms, such as $a = 1$, $b = -1$; $a = 1$, $b = -3$; $a = 1$, $b = 2$; etc., are also examined for these six alkanes, and they all fail to discriminate these five pairs (see Figure 2). So, it is not a matter of the precision of the calculations and the particular form of the mathematical manipulations (a, b, etc.), rather the duplications come from the complete correspondence of the paths, when giving the distance and the composition of the sequence of δ values along each path.

Second, an empirical formula C_8 is supplied to the structure generator of ESESOC to build up all the isomers of this molecular formula, and an exhaustive and irredundant list of total 204 structures (Charts 1−4, structures 1−204) are generated. The δ^v value of the nodes in all these graphs is the same, 4. After computing, fifteen duplications, six triple duplications, and one quadruple duplication are produced, and the duplication sets and their τ values ($a = 1$, $b = 2$) are listed in Table 2.

Last, the exhaustive and irredundant list of total 54 graphs (Chart 5, structures 205−258) are generated from the segment set containing 10 segments of >C=. The τ values of these graphs are shown in Table 3. From Table 3, it can be seen that the total topological index τ has low discrimination for this graph family. From Tables 2 and 3 and the listings of structures 1−204 and 205−208, it can be seen that the total topological index τ has high discrimination for the polycyclic graphs without multibonds and low discrimination for the polycyclic graphs with multibonds, owing to the absence of multiplicity information of the t^{ij} equation.

CONCLUSIONS

The total topological index τ is very excellent for its high discrimination of alkyl isomers and some polycyclic graphs without multibonds and has low discrimination for the polycyclic graphs with multibonds, owing to the absence of multiplicity information of Hall and Kier's method.

REFERENCES AND NOTES

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

(2) Hall, L. H.; Mohney, B.; Kier, L. B. The Electrotopological State: Structure Information at the Atomic Level for Molecular Graphs. *J. Chem. Inf. Comput. Sci.*, **1991**, *31*, 76.1

(3) Hall, L. H.; Kier, L. B. The Molecular connectivity Chi Indexes and Kappa Shape Indexes in Structure-Property Modeling. In *Reviews of Computational Chemistry*; Lipkowitz, K. R.; Boyd, D. B., Eds.; VCH Publishers: New York, 1991.

(4) Hu, C. Y.; Xu, L. A New Scheme for Assignment of Canonical Connection Table. *J. Chem. Inf. Comput. Sci.*, in press.

(5) Hu, C. Y.; Xu, L. An New Algorithm for Computer Perception of Topological Symmetry. *Anal. Chim. Acta*, in press.

(6) Hu, C. Y.; Xu, L. Studies on Expert System for the Elucidation of the Structures of Organic Compounds—Structural Generator of ESESOC-II. *Sci. China, Series B*, in press.

(7) Hu, C. Y.; Xu, L. Principles for Structure Generation of Organic Isomers from Molecular Formula. *Anal. Chim. Acta*, in press.

(8) Xu, L.; Wang, H. Y.; Su, Q. A Newly Proposed Topological Index for Discrimination of Cis/Trans Isomers. *Comput. Chem.* **1992**, *16*, 187.

(9) Xu, L. Molecular Topological Index a_N and its Extension. *J. Serb. Chem. Soc.* **1992**, *57*, 485.

(10) Yao, Y. Y.; Xu, L.; Yang, Y. Q.; Yuan, X. S. Studies on Structure-Activity Relationships of Organic Compounds—Three Topological Indexes and Their Applications. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 590.

(11) Yang, Y. Q.; Xu, L.; Hu, C. Y. Extended Adjacency Matrix Indeces and Their Applications. *J. Chem. Inf. Comput. Sci.*, in press.