## ARTICLES

# Structural Interpretation of the Topological Index. 2. The Molecular Connectivity Index, the Kappa Index, and the Atom-type E-State Index

Qian-Nan Hu,[†] Yi-Zeng Liang,*[,†] Hong Yin,[‡] Xiao-Ling Peng,[‡] and Kai-Tai Fang[‡]

Institute of Chemometrics and Intelligent Analytical Instruments, College of Chemistry and Chemical Engineering, Central South University, Changsha, 410083, P.R. China, and Statistics Research and Consultancy Centre, Hong Kong Baptist University, Hong Kong

Received January 11, 2004

The structural interpretation is extended to the topological indices describing cyclic structures. Three representatives of the topological index, such as the molecular connectivity index, the Kappa index, and the atom-type E-State index, are interpreted by mining out, through projection pursuit combining with a number theory method generating uniformly distributed directions on unit sphere, the structural features hidden in the spaces spanned by the three series of indices individually. Some interesting results, which can hardly be found by individual index, are obtained from the multidimensional spaces by several topological indices. The results support quantitatively the former studies on the topological indices, and some new insights are obtained during the analysis. The combinations of several molecular connectivity indices describe mainly three general categories of molecular structure information, which include degree of branching, size, and degree of cyclicity. The cyclicity can also be coded by the combination of chi cluster and path/cluster indices. The Kappa shape indices encode, in combination, significant information on size, the degree of cyclicity, and the degree of centralization/separation in branching. The size, branch number, and cyclicity information has also been mined out to interpret atom-type E-State indices. The structural feature such as the number of quaternary atoms is searched out to be an important factor. The results indicate that the collinearity might be a serious problem in the applications of the topological indices.

## INTRODUCTION

The topological index is commonly a descriptor hiding many finer features into one quantifier, which might be one of the reasons leading to the critique that the structural interpretation of the topological index is difficult. In the prior work,[1] the external factor variable connectivity index is interpreted by mining out the structural features hidden in the multidimensional point cloud spanned by the several indices through projection pursuit combining with the TFWW method. In the study by Katritzky et al.,[2] the retention index is analyzed to be dependent on the size, branch number, and also the branching position of molecules. In our work,[1] the three main structural features coded by the seven EFVCI indices are searched out by projection pursuit and graph center concept. To interpret the used index in QSAR/QSPR models should be helpful to interpret the built equations between the structure descriptors and the properties, and in the prior work, we have discussed the changing tendency by the variations of structural features mined out by projection pursuit. Our basic position when considering interpretation of TIs is that topological indices have an interpretation within structural chemistry.[3,4] In the present work, the methods are extended to the topological index describing cyclic structures by applying to other indices, such as the molecular connectivity index, the Kappa index, and the atom-type E-State index.

There are some studies on the structural features described by the molecular connectivity index, the Kappa index, and also the atom-type E-State index. In the monograph[5,6] by Kier and Hall, the molecular structure information described by the molecular connectivity index includes degree of branching, variable branching pattern, position and influence of heteroatoms, patterns of adjacency, and degree of cyclicity. The Kappa shape indices[7,8] encode, in combination, significant information on the degree of cyclicity and the degree of centralization/separation in branching. The atom-type E-State index,[9,10] combining both electronic and topological factors together, is interpreted by the combination with the element content, electronic organization, and the local topological state of an atom or group.

The structure information described in the references is mainly based on the analysis of an individual topological index. However, the structural features are described, in most cases, by the multidimensional description of molecular structures. In QSAR/QSPR, the tackled problems are always involved with the multidimensional description of structures. To interpret the topological index only by analysis of an individual index is not enough. To find the integration effect of all these indices upon the structural features, we need to

* Corresponding author phone: 86-731-8822841; fax: 86-731-8825637; e-mail: yizeng_liang@263.net.
[†] Central South University.
[‡] Hong Kong Baptist University.

10.1021/ci049973z CCC: $27.50 
Published on Web 04/21/2004

use a projection pursuit technique recently developed in statistics, which is supposed to be able to reveal the data structure hidden in a high-dimensional space.[11−13] In the present work, we consider the interpretation problem by analyzing some interesting projection directions in the multidimensional space spanned by several topological indices.

To find out what kinds of structural features are described by the used indices and how much an individual index contributes to a specific structural feature should be helpful in understanding the nature of the topological index. The projection pursuit,[11−13] which is concerned with "interesting" projections of high dimensional data sets to machine-pick "interesting" low-dimensional projections of a high-dimensional point cloud by numerically maximizing a certain objective function or projection index, and the TFWW[14,15] method, a quick method to generate the number-theory net (NT-net) on the unit sphere $U(U^s)$, based on the good lattice point (GLP),[16,17] are applied in the interpretation of the three topological indices. GLP is a method based on number-theory to generate uniformly distributed points in a unit space. With the help of a mathematic transformation, the points uniformly distributed in a unit space can be transferred into the number-theory net, which represents the directions uniformly distributed on the unit sphere. In the calculation procedure, we define first a projection index to measure the "interestingness" of the current projection and then rotate our projections following the directions uniformly distributed on the unit space of the NT-net. Each projection direction has an entropy value. With the rotation of projection directions on unit sphere, there will generate many entropy values. Several directions of minimum entropy values are finally selected to find some interesting structure clusters hidden in the high-dimensional space.

On one side, there are several structural features described by a topological index. On the other side, in many cases, a structural feature is often described by several indices, not by an individual index. Then, it is necessary to make two problems clear: (a) what kinds of features are coded by the several indices, that is how much a specific index contributes to a special feature, and (b) what kinds of structural features are described by a specific index.

In the work, the outlines of the methods in the prior work[1] are, first, briefly given. Second, the structural features are obtained for the different interesting projection directions. Then, the relationship between structural features and the indices is further analyzed. Finally, the structural features are applied to discover how the boiling points change with the variation of structural features by a statistical way.

## METHODS

**Outlines of the Methods.** (1) Given data descriptor matrix $X(n*m)$, generate GLP set on cubic sphere $C^s$.

(2) After getting the GLP set, the NT-net on the unit sphere $U(U^s)$ can be calculated by the TFWW algorithm.

(3) Project the matrix X onto the uniformly distributed directions.

(4) Choose some interesting directions, which hold lower entropy, an objective function computed on a projected density (or data set).

(5) Investigate the structure information on the interesting directions.

**Data Collection and Descriptors.** A well-known data set, 530 saturated hydrocarbons (without methane), is chosen from ref 18 to be the test data set. The hydrocarbons were compiled with regularity in ref 18 from methane to decanes and also from acyclic to pentacyclic hydrocarbons. The structure names (from ref 18), their boiling points, and their structural features such as size, cyclicity (calculated by the formula to compute the cyclicity used in the Balaban *J* index), and branch number are all listed in Table 1.

The indices used are as follows: ${}^0\chi_{path}$, ${}^1\chi_{path}$, ${}^2\chi_{path}$, ${}^3\chi_{path}$, ${}^3\chi_{cluster}$, and ${}^4\chi_{path}$ for the molecular connectivity index; ${}^1\kappa$, ${}^2\kappa$, ${}^3\kappa$, and $\phi$ for the Kappa index; and $S_{CH3}$, $S_{CH2}$, $S_{CH}$, and $S_C$ for the atom-type E-State index, which are calculated by the in-house software, Heuristic Queue Notation system (H.Q.N.s).[19] Atom-type E-State descriptors are derived from the combination of element content, electronic organization, and the local topological state of an atom or group. That is, the atom-type E-State index deals with both topological and mainly electronic factors. In this data set, the electronic distribution does not vary much, and the structure information is largely a topological factor.

## RESULTS AND DISCUSSION

**Regression by the Three Sets of Indices.** At first, the 530 normal boiling points (*BP*) are regressed by the individual set of topological indices with regression results as follows: $R(Chi,BP) = 0.9825$, $R$ (*atom-type E-State*,*BP*) $= 0.9824$, and $R(Kappa,BP) = 0.8761$. From the results, the three sets of indices have coded most of the information of the *BP*, although the regression might be improved further by selecting variables from a large descriptor pool, adding new variables, or using nonlinear regression. However, what kind of structural information is coded by the three sets of indices is focused in the following sections.

**Molecular Connectivity Index.** In refs 5 and 6 there are five general categories of molecular structure information intuitively described by the various Chi indices, which include the following: (1) degree of branching (emphasized in low order Chi indices); (2) variable branching pattern (emphasized in high order path Chi indices); (3) position and influence of heteroatoms (emphasized in the valence Chi indices); (4) patterns of adjacency (emphasized in the Chi cluster and path/cluster indices); and (5) degree of cyclicity (emphasized in the Chi chain indices). The statements are mainly based on the individual index. By the outlines in the Method section, the discussed structural features are quantitatively found out for different projection directions.

Projection in the direction *a1* = [−0.1869, −0.9613, 0.0052, −0.0274, −0.1994, 0.0222] by six Chi indices for 530 saturated hydrocarbons shows the size information in subplots *1a1* and *1a2* of Figure 1, in which the integral numbers on the subplot *1a1* are the number of carbon atoms (size of the molecules). The projection direction *a1* is composed of the six components corresponding to the six Chi indices. The subplots *1a1* and *1a2* are essentially the same, in which the only difference is that the *x*-axis of *1a1* and *1a2* is the sequence no. of the listed molecules and the number of carbon atoms of molecules, respectively. The variation between the different sizes is changed with an

**Table 1.** Structure Information of the 530 Hydrocarbons and Their Boiling Points[a]

| ID | name | cn | cyc | bn | V4 | bp | ID | name | cn | cyc | bn | v4 | Bp | ID | name | cn | cyc | bn | v4 | Bp |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | n2 | 2 | 0 | 0 | 0 | −88.6 | 76 | 1mc6 | 7 | 1 | 1 | 0 | 101 | 151 | C8 | 8 | 1 | 0 | 0 | 149 |
| 2 | n3 | 3 | 0 | 0 | 0 | −42.1 | 77 | c7 | 7 | 1 | 0 | 0 | 118.4 | 152 | bCprm | 8 | 2 | 0 | 0 | 129 |
| 3 | C3 | 3 | 1 | 0 | 0 | −32.8 | 78 | dcprm | 7 | 2 | 0 | 0 | 102 | 153 | bC330o | 8 | 2 | 0 | 0 | 137 |
| 4 | n4 | 4 | 0 | 0 | 0 | −0.5 | 79 | bc221h | 7 | 2 | 0 | 0 | 105.5 | 154 | bCb | 8 | 2 | 0 | 0 | 136 |
| 5 | 2mn3 | 4 | 0 | 1 | 0 | −11.7 | 80 | bc311h | 7 | 2 | 0 | 0 | 110 | 155 | bC420o | 8 | 2 | 0 | 0 | 133 |
| 6 | 1mc3 | 4 | 1 | 1 | 0 | 0.7 | 81 | bc320h | 7 | 2 | 0 | 0 | 110.5 | 156 | bC510o | 8 | 2 | 0 | 0 | 141 |
| 7 | C4 | 4 | 1 | 0 | 0 | 12.6 | 82 | bc410h | 7 | 2 | 0 | 0 | 116 | 157 | 2mbc221h | 8 | 2 | 1 | 0 | 125 |
| 8 | bcll0b | 4 | 2 | 0 | 0 | 8 | 83 | s33h | 7 | 2 | 0 | 1 | 96.5 | 158 | S34o | 8 | 2 | 0 | 1 | 128 |
| 9 | n5 | 5 | 0 | 0 | 0 | 36 | 84 | s24h | 7 | 2 | 0 | 1 | 98.5 | 159 | 7mbc221h | 8 | 2 | 1 | 0 | 128 |
| 10 | 2mn4 | 5 | 0 | 1 | 0 | 27.8 | 85 | 2mbc310hx | 7 | 2 | 1 | 0 | 100 | 160 | 2mbc320h | 8 | 2 | 1 | 0 | 130.5 |
| 11 | 22mn3 | 5 | 0 | 2 | 1 | 9.5 | 86 | 6mbc310hx | 7 | 2 | 1 | 0 | 103 | 161 | S25o | 8 | 2 | 0 | 1 | 125 |
| 12 | 1ec3 | 5 | 1 | 1 | 0 | 35.9 | 87 | mbc211hx | 7 | 2 | 1 | 1 | 81.5 | 162 | 1mbc221h | 8 | 2 | 1 | 1 | 117 |
| 13 | 12mc3 | 5 | 1 | 2 | 0 | 32.6 | 88 | mbc310hx | 7 | 2 | 1 | 1 | 92 | 163 | 7mbc410h | 8 | 2 | 1 | 0 | 138 |
| 14 | 11mc3 | 5 | 1 | 2 | 1 | 20.6 | 89 | 13mbc111p | 7 | 2 | 2 | 2 | 71.5 | 164 | 1mbc410h | 8 | 2 | 1 | 1 | 125 |
| 15 | 1mc4 | 5 | 1 | 1 | 0 | 36.3 | 90 | 14mbc210p | 7 | 2 | 2 | 2 | 74 | 165 | 33mbc310hx | 8 | 2 | 2 | 1 | 115 |
| 16 | c5 | 5 | 1 | 0 | 0 | 49.3 | 91 | 11ms22p | 7 | 2 | 2 | 2 | 78 | 166 | 14mbc211hx | 8 | 2 | 2 | 2 | 91 |
| 17 | bclllp | 5 | 2 | 0 | 0 | 36 | 92 | 122mbcb | 7 | 2 | 3 | 2 | 84 | 167 | 66mbC310hX | 8 | 2 | 2 | 1 | 126.1 |
| 18 | bc210p | 5 | 2 | 0 | 0 | 46 | 93 | tc410024h | 7 | 3 | 0 | 0 | 105 | 168 | 2244mbcb | 8 | 2 | 4 | 2 | 104 |
| 19 | s22p | 5 | 2 | 0 | 1 | 39 | 94 | tc311024h | 7 | 3 | 0 | 0 | 107 | 169 | 1223mbcb | 8 | 2 | 4 | 3 | 105 |
| 20 | mbcllob | 5 | 2 | 1 | 1 | 33.5 | 95 | tc221026h | 7 | 3 | 0 | 0 | 106 | 170 | tc510035o | 8 | 3 | 0 | 0 | 142 |
| 21 | n6 | 6 | 0 | 0 | 0 | 68.7 | 96 | tc410027h | 7 | 3 | 0 | 0 | 110 | 171 | tc510024o | 8 | 3 | 0 | 0 | 149 |
| 22 | 2mn5 | 6 | 0 | 1 | 0 | 60.3 | 97 | tc410013h | 7 | 3 | 0 | 1 | 107.5 | 172 | tc3210o | 8 | 3 | 0 | 0 | 136 |
| 23 | 3mn5 | 6 | 0 | 1 | 0 | 63.3 | 98 | tec320h | 7 | 4 | 0 | 0 | 108.5 | 173 | tc3300o | 8 | 3 | 0 | 0 | 125 |
| 24 | 23mn4 | 6 | 0 | 2 | 0 | 58 | 99 | tec410h | 7 | 4 | 0 | 0 | 104 | 174 | 3mtc2210h | 8 | 3 | 1 | 0 | 120.5 |
| 25 | 22mn4 | 6 | 0 | 2 | 1 | 49.7 | 100 | n8 | 8 | 0 | 0 | 0 | 125.7 | 175 | ds2121o | 8 | 3 | 0 | 2 | 103 |
| 26 | 1pc3 | 6 | 1 | 1 | 0 | 69 | 101 | 2mn7 | 8 | 0 | 1 | 0 | 117.6 | 176 | 1mtc2210h | 8 | 3 | 1 | 1 | 111 |
| 27 | 1ipc3 | 6 | 1 | 2 | 0 | 58.3 | 102 | 3mn7 | 8 | 0 | 1 | 0 | 118.9 | 177 | ds2022o | 8 | 3 | 0 | 2 | 115 |
| 28 | 1e2mc3 | 6 | 1 | 2 | 0 | 63 | 103 | 4mn7 | 8 | 0 | 1 | 0 | 117.7 | 178 | tec330o | 8 | 4 | 0 | 0 | 137.5 |
| 29 | 1elmc3 | 6 | 1 | 2 | 1 | 57 | 104 | 25mn6 | 8 | 0 | 2 | 0 | 109.1 | 179 | n9 | 9 | 0 | 0 | 0 | 150.8 |
| 30 | 123mc3 | 6 | 1 | 3 | 0 | 63 | 105 | 3en6 | 8 | 0 | 1 | 0 | 118.5 | 180 | 2mn8 | 9 | 0 | 1 | 0 | 142.8 |
| 31 | 112mc3 | 6 | 1 | 3 | 1 | 52.6 | 106 | 24mn6 | 8 | 0 | 2 | 0 | 109.4 | 181 | 3mn8 | 9 | 0 | 1 | 0 | 144 |
| 32 | 1ec4 | 6 | 1 | 1 | 0 | 70.7 | 107 | 23mn6 | 8 | 0 | 2 | 0 | 115.6 | 182 | 4mn8 | 9 | 0 | 1 | 0 | 142.4 |
| 33 | 13mc4 | 6 | 1 | 2 | 0 | 59 | 108 | 34mn6 | 8 | 0 | 2 | 0 | 117.7 | 183 | 26mn7 | 9 | 0 | 2 | 0 | 134 |
| 34 | 12mc4 | 6 | 1 | 2 | 0 | 62 | 109 | 22mn6 | 8 | 0 | 2 | 1 | 106.8 | 184 | 3en7 | 9 | 0 | 1 | 0 | 143 |
| 35 | 11mc4 | 6 | 1 | 2 | 1 | 53.6 | 110 | 3e2mn5 | 8 | 0 | 2 | 0 | 115.6 | 185 | 4en7 | 9 | 0 | 1 | 0 | 142.1 |
| 36 | 1mc5 | 6 | 1 | 1 | 0 | 71.8 | 111 | 234mn5 | 8 | 0 | 3 | 0 | 113.5 | 186 | 25mn7 | 9 | 0 | 2 | 0 | 136 |
| 37 | c6 | 6 | 1 | 0 | 0 | 80.7 | 112 | 33mn6 | 8 | 0 | 2 | 1 | 112 | 187 | 24mn7 | 9 | 0 | 2 | 0 | 133.5 |
| 38 | bc211hx | 6 | 2 | 0 | 0 | 71 | 113 | 224mn5 | 8 | 0 | 3 | 1 | 99.2 | 188 | 23mn7 | 9 | 0 | 2 | 0 | 140.5 |
| 39 | bcpr | 6 | 2 | 0 | 0 | 76 | 114 | 3e3mn5 | 8 | 0 | 2 | 1 | 118.2 | 189 | 35mn7 | 9 | 0 | 2 | 0 | 136 |
| 40 | bc220hx | 6 | 2 | 0 | 0 | 83 | 115 | 223mn5 | 8 | 0 | 3 | 1 | 109.8 | 190 | 2m4en6 | 9 | 0 | 2 | 0 | 133.8 |
| 41 | bc310hx | 6 | 2 | 0 | 0 | 81 | 116 | 233mn5 | 8 | 0 | 3 | 1 | 114.8 | 191 | 22mn7 | 9 | 0 | 2 | 1 | 132.7 |
| 42 | s23hx | 6 | 2 | 0 | 1 | 69.5 | 117 | 2233mn4 | 8 | 0 | 4 | 2 | 106.5 | 192 | 34mn7 | 9 | 0 | 2 | 0 | 140.6 |
| 43 | mbc210p | 6 | 2 | 1 | 1 | 60.5 | 118 | 1pec3 | 8 | 1 | 1 | 0 | 128 | 193 | 2m3en6 | 9 | 0 | 2 | 0 | 138 |
| 44 | 13mbcb | 6 | 2 | 2 | 2 | 55 | 119 | 1spec3 | 8 | 1 | 2 | 0 | 117.7 | 194 | 235mn6 | 9 | 0 | 3 | 0 | 131.3 |
| 45 | n7 | 7 | 0 | 0 | 0 | 98.5 | 120 | b2mc3 | 8 | 1 | 2 | 0 | 124 | 195 | 3m4en6 | 9 | 0 | 2 | 0 | 140.4 |
| 46 | 2mn6 | 7 | 0 | 1 | 0 | 90 | 121 | 1nepec3 | 8 | 1 | 3 | 1 | 106 | 196 | 225mn6 | 9 | 0 | 3 | 1 | 124 |
| 47 | 3mn6 | 7 | 0 | 1 | 0 | 92 | 122 | 5msbc3 | 8 | 1 | 3 | 0 | 115.5 | 197 | 33mn7 | 9 | 0 | 2 | 1 | 137.3 |
| 48 | 3en5 | 7 | 0 | 1 | 0 | 93.5 | 123 | 1e2pc3 | 8 | 1 | 2 | 0 | 108 | 198 | 44mn7 | 9 | 0 | 2 | 1 | 135.2 |
| 49 | 24mn5 | 7 | 0 | 2 | 0 | 80.5 | 124 | ib2mc3 | 8 | 1 | 3 | 0 | 110 | 199 | 234mn6 | 9 | 0 | 3 | 0 | 139 |
| 50 | 23mn5 | 7 | 0 | 2 | 0 | 89.8 | 125 | 11m2pc3 | 8 | 1 | 3 | 1 | 105.9 | 200 | 224mn6 | 9 | 0 | 3 | 1 | 126.5 |
| 51 | 22mn5 | 7 | 0 | 2 | 1 | 79.2 | 126 | 1m12ec3 | 8 | 1 | 3 | 1 | 108.9 | 201 | 24m3en5 | 9 | 0 | 3 | 0 | 136.7 |
| 52 | 33mn5 | 7 | 0 | 2 | 1 | 86.1 | 127 | 11m2ipc3 | 8 | 1 | 4 | 1 | 94.4 | 202 | 3m3en6 | 9 | 0 | 2 | 1 | 140.6 |
| 53 | 223mn4 | 7 | 0 | 3 | 1 | 80.9 | 128 | 112m2ec3 | 8 | 1 | 4 | 2 | 104.5 | 203 | 244mn6 | 9 | 0 | 3 | 1 | 130.7 |
| 54 | 1bc3 | 7 | 1 | 1 | 0 | 98 | 129 | 11223mc3 | 8 | 1 | 5 | 2 | 100.5 | 204 | 223mn6 | 9 | 0 | 3 | 1 | 133.6 |
| 55 | 1sbc3 | 7 | 1 | 2 | 0 | 90.3 | 130 | 1ibc4 | 8 | 1 | 2 | 0 | 120.1 | 205 | 33en5 | 9 | 0 | 2 | 1 | 145 |
| 56 | 1m2pc3 | 7 | 1 | 2 | 0 | 93 | 131 | p3mc4 | 8 | 1 | 2 | 0 | 117.4 | 206 | 233mn6 | 9 | 0 | 3 | 1 | 137.7 |
| 57 | 12ec3 | 7 | 1 | 2 | 0 | 90 | 132 | 1sbc4 | 8 | 1 | 2 | 0 | 123 | 207 | 22m3en5 | 9 | 0 | 3 | 1 | 133.8 |
| 58 | 1mlpc3 | 7 | 1 | 2 | 1 | 84.9 | 133 | 12ec4 | 8 | 1 | 2 | 0 | 119 | 208 | 334mn6 | 9 | 0 | 3 | 1 | 140.5 |
| 59 | 1m2ipc3 | 7 | 1 | 3 | 0 | 81.1 | 134 | 1234mc4 | 8 | 1 | 4 | 0 | 114.5 | 209 | 23m3en5 | 9 | 0 | 3 | 1 | 142 |
| 60 | 1tbc3 | 7 | 1 | 3 | 1 | 80.5 | 135 | 1133mc4 | 8 | 1 | 4 | 2 | 86 | 210 | 2244mn5 | 9 | 0 | 4 | 2 | 122.3 |
| 61 | 11ec3 | 7 | 1 | 2 | 1 | 88.6 | 136 | 1pc5 | 8 | 1 | 1 | 0 | 131 | 211 | 2234mn5 | 9 | 0 | 4 | 1 | 133 |
| 62 | 1e23mc3 | 7 | 1 | 3 | 0 | 91 | 137 | 1ipc5 | 8 | 1 | 2 | 0 | 126.4 | 212 | 2334mn5 | 9 | 0 | 4 | 1 | 141.5 |
| 63 | 1mlipc3 | 7 | 1 | 3 | 1 | 81.5 | 138 | 1e3mc5 | 8 | 1 | 2 | 0 | 121 | 213 | 2233mn5 | 9 | 0 | 4 | 2 | 140.2 |
| 64 | 11m2ec3 | 7 | 1 | 3 | 1 | 79.1 | 139 | 1e2mc5 | 8 | 1 | 2 | 0 | 124.7 | 214 | 1hxc3 | 9 | 1 | 1 | 0 | 149 |
| 65 | 12mlec3 | 7 | 1 | 3 | 1 | 85.2 | 140 | 124mc5 | 8 | 1 | 3 | 0 | 115 | 215 | 1ShXC3 | 9 | 1 | 2 | 0 | 143 |
| 66 | 1123mc3 | 7 | 1 | 4 | 1 | 78 | 141 | 1elmc5 | 8 | 1 | 2 | 1 | 121.5 | 216 | 1m2pec3 | 9 | 1 | 2 | 0 | 153 |
| 67 | 1122mc3 | 7 | 1 | 4 | 2 | 76 | 142 | 123mc5 | 8 | 1 | 3 | 0 | 117 | 217 | 12pC3 | 9 | 1 | 2 | 0 | 142 |
| 68 | 1pC4 | 7 | 1 | 1 | 0 | 100.7 | 143 | 113mc5 | 8 | 1 | 3 | 1 | 104.9 | 218 | 1elbc3 | 9 | 1 | 2 | 1 | 140.2 |
| 69 | 1ipc4 | 7 | 1 | 2 | 0 | 92.7 | 144 | 112mc5 | 8 | 1 | 3 | 1 | 114 | 219 | nepelmc3 | 9 | 1 | 4 | 2 | 125 |
| 70 | 1e3mc4 | 7 | 1 | 2 | 0 | 89.5 | 145 | 1ec6 | 8 | 1 | 1 | 0 | 131.8 | 220 | 11m2ibc3 | 9 | 1 | 4 | 1 | 125.5 |
| 71 | 1e2mc4 | 7 | 1 | 2 | 0 | 94 | 146 | 14mc6 | 8 | 1 | 2 | 0 | 121.8 | 221 | tb22mc3 | 9 | 1 | 5 | 2 | 121 |
| 72 | 1ec5 | 7 | 1 | 1 | 0 | 103.5 | 147 | 13mc6 | 8 | 1 | 2 | 0 | 122.3 | 222 | 12m12ec3 | 9 | 1 | 4 | 2 | 130.8 |
| 73 | 13mc5 | 7 | 1 | 2 | 0 | 91.3 | 148 | 12mc6 | 8 | 1 | 2 | 0 | 126.6 | 223 | 112233mc3 | 9 | 1 | 6 | 3 | 124.5 |
| 74 | 12mc5 | 7 | 1 | 2 | 0 | 95.6 | 149 | 11mc6 | 8 | 1 | 2 | 1 | 119.5 | 224 | 3pec4 | 9 | 1 | 2 | 0 | 148.7 |
| 75 | 11mc5 | 7 | 1 | 2 | 1 | 87.9 | 150 | 1mc7 | 8 | 1 | 1 | 0 | 134 | 225 | 1bC5 | 9 | 1 | 1 | 0 | 156.6 |

**Table 1** (Continued)

| ID | name | cn | cyc | bn | V4 | bp | ID | name | cn | cyc | bn | v4 | Bp | ID | name | cn | cyc | bn | v4 | Bp |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 226 | 1ibc5 | 9 | 1 | 2 | 0 | 148 | 301 | ebc410h | 9 | 2 | 1 | 1 | 150.5 | 376 | 22m3en6 | 10 | 0 | 3 | 1 | 159 |
| 227 | 1m3pc5 | 9 | 1 | 2 | 0 | 148.3 | 302 | 15ms33h | 9 | 2 | 2 | 1 | 132.2 | 377 | 2235mn6 | 10 | 0 | 4 | 1 | 148.7 |
| 228 | 1SbC5 | 9 | 1 | 2 | 0 | 154.3 | 303 | 77mbc410h | 9 | 2 | 2 | 1 | 154.5 | 378 | 33en6 | 10 | 0 | 2 | 1 | 166.3 |
| 229 | 13ec5 | 9 | 1 | 2 | 0 | 148.2 | 304 | 266mbc310hx | 9 | 2 | 3 | 1 | 141 | 379 | 24m3ipn5 | 10 | 0 | 4 | 0 | 157 |
| 230 | 1m2pc5 | 9 | 1 | 2 | 0 | 149.5 | 305 | 155mbc211hX | 9 | 2 | 3 | 2 | 131 | 380 | 334mn7 | 10 | 0 | 3 | 1 | 164 |
| 231 | 12ec5 | 9 | 1 | 2 | 0 | 150.5 | 306 | 1m5ebc310hx | 9 | 2 | 2 | 2 | 135 | 381 | 344mn7 | 10 | 0 | 3 | 1 | 164 |
| 232 | 1m3ipc5 | 9 | 1 | 3 | 0 | 141 | 307 | 135mbc310hx | 9 | 2 | 3 | 2 | 127.2 | 382 | 2335mn6 | 10 | 0 | 4 | 1 | 153 |
| 233 | 1m2ipc5 | 9 | 1 | 3 | 0 | 145 | 308 | tc331037n | 9 | 3 | 0 | 0 | 164.6 | 383 | 33m4en6 | 10 | 0 | 3 | 1 | 165 |
| 234 | 1mlpc5 | 9 | 1 | 2 | 1 | 145 | 309 | tC421037n | 9 | 3 | 0 | 0 | 162.5 | 384 | 23m3en6 | 10 | 0 | 3 | 1 | 169 |
| 235 | 12m4ec5 | 9 | 1 | 3 | 0 | 142.5 | 310 | tC421024n | 9 | 3 | 0 | 0 | 166 | 385 | 2244mn6 | 10 | 0 | 4 | 2 | 153 |
| 236 | 13m4ec5 | 9 | 1 | 3 | 0 | 138 | 311 | tc430037n | 9 | 3 | 0 | 0 | 161.4 | 386 | 2234mn6 | 10 | 0 | 4 | 1 | 155 |
| 237 | 12m3ec5 | 9 | 1 | 3 | 0 | 147 | 312 | dS2122n | 9 | 3 | 0 | 2 | 142.5 | 387 | 34m3en6 | 10 | 0 | 3 | 1 | 170 |
| 238 | 13m2ec5 | 9 | 1 | 3 | 0 | 151 | 313 | 11Cprc3 | 9 | 3 | 0 | 1 | 147.8 | 388 | 224m3en5 | 10 | 0 | 4 | 1 | 155.3 |
| 239 | 1tbC5 | 9 | 1 | 3 | 1 | 145 | 314 | Scprnorb | 9 | 3 | 0 | 1 | 140.7 | 389 | 2344mn6 | 10 | 0 | 4 | 1 | 162 |
| 240 | 11ec5 | 9 | 1 | 2 | 1 | 151 | 315 | 3mtc3210o | 9 | 3 | 1 | 0 | 161 | 390 | 2m33en5 | 10 | 0 | 3 | 1 | 174 |
| 241 | 11m3ec5 | 9 | 1 | 3 | 1 | 133 | 316 | dS2023n | 9 | 3 | 0 | 2 | 146 | 391 | 2334mn6 | 10 | 0 | 4 | 1 | 164 |
| 242 | 13mlec5 | 9 | 1 | 3 | 1 | 135.5 | 317 | etc2210h | 9 | 3 | 1 | 1 | 137.5 | 392 | 23m3ipn5 | 10 | 0 | 4 | 1 | 169.4 |
| 243 | 1m1ipc5 | 9 | 1 | 3 | 1 | 143 | 318 | 33mtc2210h | 9 | 3 | 2 | 1 | 137.5 | 393 | 2233mn6 | 10 | 0 | 4 | 2 | 160 |
| 244 | 11m2ec5 | 9 | 1 | 3 | 1 | 138 | 319 | 17mtc2210h | 9 | 3 | 2 | 1 | 131 | 394 | 22344mn5 | 10 | 0 | 5 | 2 | 159.3 |
| 245 | 1234mc5 | 9 | 1 | 4 | 0 | 135.9 | 320 | 12mtc2210h | 9 | 3 | 2 | 2 | 128 | 395 | 3344mn6 | 10 | 0 | 4 | 2 | 170.5 |
| 246 | 12mlec5 | 9 | 1 | 3 | 1 | 142.5 | 321 | tec33100n | 9 | 4 | 0 | 0 | 168.3 | 396 | 223m3en5 | 10 | 0 | 4 | 2 | 168 |
| 247 | 1134mc5 | 9 | 1 | 4 | 1 | 127 | 322 | tec43000n | 9 | 4 | 0 | 0 | 153 | 397 | 22334mn5 | 10 | 0 | 5 | 2 | 166 |
| 248 | 1124mc5 | 9 | 1 | 4 | 1 | 129.5 | 323 | n10 | 10 | 0 | 0 | 0 | 174.1 | 398 | 1ihxlmc3 | 10 | 1 | 3 | 1 | 155.1 |
| 249 | 1123mc5 | 9 | 1 | 4 | 1 | 134 | 324 | 2mn9 | 10 | 0 | 1 | 0 | 166.8 | 399 | 1ip2ibc3 | 10 | 1 | 4 | 0 | 148.3 |
| 250 | 1133mc5 | 9 | 1 | 4 | 2 | 118.2 | 325 | 3mn9 | 10 | 0 | 1 | 0 | 167.8 | 400 | 11m2pec3 | 10 | 1 | 3 | 1 | 157.8 |
| 251 | 1223mc5 | 9 | 1 | 4 | 1 | 138 | 326 | 4mn9 | 10 | 0 | 1 | 0 | 165.7 | 401 | 12plmc3 | 10 | 1 | 3 | 1 | 153.5 |
| 252 | 1122mc5 | 9 | 1 | 4 | 2 | 135 | 327 | 27mn8 | 10 | 0 | 2 | 0 | 160 | 402 | 11m2tpec3 | 10 | 1 | 5 | 2 | 146.7 |
| 253 | 1pC6 | 9 | 1 | 1 | 0 | 156.7 | 328 | 5mn9 | 10 | 0 | 1 | 0 | 165.1 | 403 | 112m3tbc3 | 10 | 1 | 6 | 2 | 146 |
| 254 | 1ipc6 | 9 | 1 | 2 | 0 | 154.8 | 329 | 3en8 | 10 | 0 | 1 | 0 | 168 | 404 | 12ipC4 | 10 | 1 | 4 | 0 | 158 |
| 255 | 1m4ec6 | 9 | 1 | 2 | 0 | 150.8 | 330 | 26mn8 | 10 | 0 | 2 | 0 | 158.5 | 405 | 12m34ec4 | 10 | 1 | 4 | 0 | 155.5 |
| 256 | 1m3ec6 | 9 | 1 | 2 | 0 | 150 | 331 | 4en8 | 10 | 0 | 1 | 0 | 164 | 406 | 13e22mc4 | 10 | 1 | 4 | 1 | 154 |
| 257 | 1m2ec6 | 9 | 1 | 2 | 0 | 154.3 | 332 | 25mn8 | 10 | 0 | 2 | 0 | 157 | 407 | 112m3ipc4 | 10 | 1 | 5 | 1 | 145.5 |
| 258 | 135mc6 | 9 | 1 | 3 | 0 | 139.5 | 333 | 4pn7 | 10 | 0 | 1 | 0 | 161.8 | 408 | 1pec5 | 10 | 1 | 1 | 0 | 180 |
| 259 | 124mc6 | 9 | 1 | 3 | 0 | 144.8 | 334 | 24mn8 | 10 | 0 | 2 | 0 | 153 | 409 | 1ipec5 | 10 | 1 | 2 | 0 | 172 |
| 260 | 123mc6 | 9 | 1 | 3 | 0 | 149.4 | 335 | 36mn8 | 10 | 0 | 2 | 0 | 159.5 | 410 | 1m3bc5 | 10 | 1 | 2 | 0 | 170.7 |
| 261 | 1mlec6 | 9 | 1 | 2 | 1 | 152 | 336 | 23mn8 | 10 | 0 | 2 | 0 | 164 | 411 | 1Spec5 | 10 | 1 | 2 | 0 | 176.5 |
| 262 | 114mc6 | 9 | 1 | 3 | 1 | 136 | 337 | 2m5en7 | 10 | 0 | 2 | 0 | 159.7 | 412 | 1m2bc5 | 10 | 1 | 2 | 0 | 172 |
| 263 | 113mc6 | 9 | 1 | 3 | 1 | 136.6 | 338 | 35mn8 | 10 | 0 | 2 | 0 | 159 | 413 | 1m3ibc5 | 10 | 1 | 3 | 0 | 153 |
| 264 | 112mc6 | 9 | 1 | 3 | 1 | 145.1 | 339 | 22mn8 | 10 | 0 | 2 | 1 | 155 | 414 | 3pec5 | 10 | 1 | 2 | 0 | 175.7 |
| 265 | 1ec7 | 9 | 1 | 1 | 0 | 163.7 | 340 | 2m4en7 | 10 | 0 | 2 | 0 | 160 | 415 | 1nepec5 | 10 | 1 | 3 | 1 | 173.9 |
| 266 | 14mc7 | 9 | 1 | 2 | 0 | 153.8 | 341 | 246mn7 | 10 | 0 | 3 | 0 | 145 | 416 | 6mibc5 | 10 | 1 | 3 | 0 | 173.5 |
| 267 | 13mc7 | 9 | 1 | 2 | 0 | 151 | 342 | 34mn8 | 10 | 0 | 2 | 0 | 166 | 417 | 1e2ipc5 | 10 | 1 | 3 | 0 | 160 |
| 268 | 12mc7 | 9 | 1 | 2 | 0 | 157 | 343 | 3m5en7 | 10 | 0 | 2 | 0 | 160 | 418 | 13e2mc5 | 10 | 1 | 3 | 0 | 172 |
| 269 | 11mc7 | 9 | 1 | 2 | 1 | 150 | 344 | 236mn7 | 10 | 0 | 3 | 0 | 155.7 | 419 | 1tpec5 | 10 | 1 | 3 | 1 | 173.9 |
| 270 | 1mc8 | 9 | 1 | 1 | 0 | 168.2 | 345 | 2m3en7 | 10 | 0 | 2 | 0 | 166 | 420 | 12m3ipc5 | 10 | 1 | 4 | 0 | 160.5 |
| 271 | C9 | 9 | 1 | 0 | 0 | 175 | 346 | 45mn8 | 10 | 0 | 2 | 0 | 162.4 | 421 | 11m3ipc5 | 10 | 1 | 4 | 1 | 148.5 |
| 272 | bC331n | 9 | 2 | 0 | 0 | 169 | 347 | 4ipn7 | 10 | 0 | 2 | 0 | 159.5 | 422 | 123m4ec5 | 10 | 1 | 4 | 0 | 172 |
| 273 | dc4m | 9 | 2 | 0 | 0 | 161.8 | 348 | 226mn7 | 10 | 0 | 3 | 1 | 148.2 | 423 | 124m3ec5 | 10 | 1 | 4 | 0 | 171 |
| 274 | bC421n | 9 | 2 | 0 | 0 | 163 | 349 | 4m3en7 | 10 | 0 | 2 | 0 | 167 | 424 | 11m2ipc5 | 10 | 1 | 4 | 1 | 163 |
| 275 | bc430n | 9 | 2 | 0 | 0 | 163 | 350 | 235mn7 | 10 | 0 | 3 | 0 | 159.7 | 425 | 12345mc5 | 10 | 1 | 5 | 0 | 164 |
| 276 | bc520n | 9 | 2 | 0 | 0 | 164 | 351 | 3m4en7 | 10 | 0 | 2 | 0 | 167 | 426 | 11334mc5 | 10 | 1 | 5 | 2 | 141.5 |
| 277 | 1CpC6 | 9 | 2 | 0 | 0 | 157.8 | 352 | 33mn8 | 10 | 0 | 2 | 1 | 161.2 | 427 | 1bC6 | 10 | 1 | 1 | 0 | 180.9 |
| 278 | bc610n | 9 | 2 | 0 | 0 | 168 | 353 | 245mn7 | 10 | 0 | 3 | 0 | 157 | 428 | 1ibc6 | 10 | 1 | 2 | 0 | 171.3 |
| 279 | 2mbc222o | 9 | 2 | 1 | 0 | 159 | 354 | 225mn7 | 10 | 0 | 3 | 1 | 147 | 429 | m4pc6 | 10 | 1 | 2 | 0 | 173.4 |
| 280 | 2mbc321o | 9 | 2 | 1 | 0 | 155 | 355 | 25m3en6 | 10 | 0 | 3 | 0 | 157 | 430 | m3pc6 | 10 | 1 | 2 | 0 | 169 |
| 281 | 6mbc321o | 9 | 2 | 1 | 0 | 154 | 356 | 44mn8 | 10 | 0 | 2 | 1 | 160 | 431 | 1SbC6 | 10 | 1 | 2 | 0 | 179.3 |
| 282 | 3mbc330o | 9 | 2 | 1 | 0 | 152.5 | 357 | 34en6 | 10 | 0 | 2 | 0 | 162 | 432 | 14ec6 | 10 | 1 | 2 | 0 | 175.5 |
| 283 | S44n | 9 | 2 | 0 | 1 | 157 | 358 | 224mn7 | 10 | 0 | 3 | 1 | 147.7 | 433 | 13ec6 | 10 | 1 | 2 | 0 | 172 |
| 284 | S35n | 9 | 2 | 0 | 1 | 159 | 359 | 234mn7 | 10 | 0 | 3 | 0 | 163 | 434 | m2pc6 | 10 | 1 | 2 | 0 | 174.5 |
| 285 | 2mbc330o | 9 | 2 | 1 | 0 | 149 | 360 | 255mn7 | 10 | 0 | 3 | 1 | 152.8 | 435 | 1m4ipc6 | 10 | 1 | 3 | 0 | 170 |
| 286 | S26n | 9 | 2 | 0 | 1 | 153 | 361 | 2m3ipn6 | 10 | 0 | 3 | 0 | 163 | 436 | 12ec6 | 10 | 1 | 2 | 0 | 176 |
| 287 | mbc222o | 9 | 2 | 1 | 1 | 145 | 362 | 345mn7 | 10 | 0 | 3 | 0 | 164 | 437 | 1m3ipc6 | 10 | 1 | 3 | 0 | 167 |
| 288 | 2ebc221h | 9 | 2 | 1 | 0 | 152 | 363 | 22m4en6 | 10 | 0 | 3 | 1 | 147 | 438 | 1e35mc6 | 10 | 1 | 3 | 0 | 168.5 |
| 289 | mbc321o | 9 | 2 | 1 | 1 | 148.2 | 364 | 3m3en7 | 10 | 0 | 2 | 1 | 163.8 | 439 | 1m2ipc6 | 10 | 1 | 3 | 0 | 171 |
| 290 | 23mbc221h | 9 | 2 | 2 | 0 | 151 | 365 | 23m4en6 | 10 | 0 | 3 | 0 | 164 | 440 | 1mlpc6 | 10 | 1 | 2 | 1 | 174.3 |
| 291 | ebc221h | 9 | 2 | 1 | 1 | 146.7 | 366 | 24m3en6 | 10 | 0 | 3 | 0 | 164 | 441 | 12m3ec6 | 10 | 1 | 3 | 0 | 175 |
| 292 | 22mbc311h | 9 | 2 | 2 | 1 | 151.5 | 367 | 244mn7 | 10 | 0 | 3 | 1 | 152 | 442 | 1tbC6 | 10 | 1 | 3 | 1 | 171.5 |
| 293 | 4ms25o | 9 | 2 | 1 | 1 | 149 | 368 | 223mn7 | 10 | 0 | 3 | 1 | 158 | 443 | 11ec6 | 10 | 1 | 2 | 1 | 179.5 |
| 294 | mbc510o | 9 | 2 | 1 | 1 | 151 | 369 | 335mn7 | 10 | 0 | 3 | 1 | 155.7 | 444 | 14mlec6 | 10 | 1 | 3 | 1 | 168 |
| 295 | 22mbc221h | 9 | 2 | 2 | 1 | 143 | 370 | 4m4en7 | 10 | 0 | 2 | 1 | 166 | 445 | 1245mc6 | 10 | 1 | 4 | 0 | 167 |
| 296 | 12mbc221h | 9 | 2 | 2 | 1 | 138.8 | 371 | 2245mn6 | 10 | 0 | 4 | 1 | 147.9 | 446 | 13mlec6 | 10 | 1 | 3 | 1 | 166.6 |
| 297 | 1ms25o | 9 | 2 | 1 | 1 | 146.5 | 372 | 2255mn6 | 10 | 0 | 4 | 2 | 137 | 447 | 1235mc6 | 10 | 1 | 4 | 0 | 166.5 |
| 298 | 17mbc221h | 9 | 2 | 2 | 1 | 142 | 373 | 2345mn6 | 10 | 0 | 4 | 0 | 158 | 448 | 1mlipc6 | 10 | 1 | 3 | 1 | 177.5 |
| 299 | 77mbc221h | 9 | 2 | 2 | 1 | 148.2 | 374 | 233mn7 | 10 | 0 | 3 | 1 | 160 | 449 | 1234mc6 | 10 | 1 | 4 | 0 | 172.5 |
| 300 | 66mbc311h | 9 | 2 | 2 | 1 | 150 | 375 | 24m4en6 | 10 | 0 | 3 | 1 | 158 | 450 | 1135mc6 | 10 | 1 | 4 | 1 | 153 |

**Table 1** (Continued)

| ID | name | cn | cyc | bn | V4 | bp | ID | name | cn | cyc | bn | v4 | Bp | ID | name | cn | cyc | bn | v4 | Bp |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 451 | 1134mc6 | 10 | 1 | 4 | 1 | 160.3 | 478 | 8mbc430n | 10 | 2 | 1 | 0 | 173.5 | 505 | 123mbc221h | 10 | 2 | 3 | 1 | 159.5 |
| 452 | 1224mc6 | 10 | 1 | 4 | 1 | 158 | 479 | 9mbc331n | 10 | 2 | 1 | 0 | 189.9 | 506 | 277mbc221h | 10 | 2 | 3 | 1 | 163 |
| 453 | 1144mc6 | 10 | 1 | 4 | 2 | 153 | 480 | S45d | 10 | 2 | 0 | 1 | 185 | 507 | 133mbc221h | 10 | 2 | 3 | 2 | 150 |
| 454 | 1133mc6 | 10 | 1 | 4 | 2 | 155 | 481 | 2mbc430n | 10 | 2 | 1 | 0 | 182 | 508 | 266mbc311h | 10 | 2 | 3 | 1 | 168 |
| 455 | 1123mc6 | 10 | 1 | 4 | 1 | 167 | 482 | s36d | 10 | 2 | 0 | 1 | 183 | 509 | 147mbc221h | 10 | 2 | 3 | 2 | 153 |
| 456 | 1122mc6 | 10 | 1 | 4 | 2 | 161.5 | 483 | 2ebc222o | 10 | 2 | 1 | 0 | 183 | 510 | 377mbc410h | 10 | 2 | 3 | 1 | 168.5 |
| 457 | 1pc7 | 10 | 1 | 1 | 0 | 182.8 | 484 | 7mbc430n | 10 | 2 | 1 | 0 | 182 | 511 | 177mbc221h | 10 | 2 | 3 | 2 | 160 |
| 458 | 135mc7 | 10 | 1 | 3 | 0 | 164.2 | 485 | 1mbc331n | 10 | 2 | 1 | 1 | 178 | 512 | 11ms25o | 10 | 2 | 2 | 2 | 152 |
| 459 | 125mc7 | 10 | 1 | 3 | 0 | 173 | 486 | 2ebc330o | 10 | 2 | 1 | 0 | 174 | 513 | 1ip4mbc310hx | 10 | 2 | 3 | 1 | 157.5 |
| 460 | 124mc7 | 10 | 1 | 3 | 0 | 170.7 | 487 | 23mbc321o | 10 | 2 | 2 | 0 | 174.5 | 514 | 1p5mbc310hx | 10 | 2 | 2 | 2 | 158.5 |
| 461 | 123mc7 | 10 | 1 | 3 | 0 | 177.5 | 488 | 26mbc222o | 10 | 2 | 2 | 0 | 172.5 | 515 | 15m3ebc310hx | 10 | 2 | 3 | 2 | 159.5 |
| 462 | 113mc7 | 10 | 1 | 3 | 1 | 161 | 489 | 37mbc330o | 10 | 2 | 2 | 0 | 166 | 516 | 1ip5mbc310hx | 10 | 2 | 3 | 2 | 152 |
| 463 | 112mc7 | 10 | 1 | 3 | 1 | 174 | 490 | 26mbc321o | 10 | 2 | 2 | 0 | 164.5 | 517 | 3cpbc410h | 10 | 3 | 0 | 0 | 175.5 |
| 464 | 1ec8 | 10 | 1 | 1 | 0 | 191.4 | 491 | ebc222o | 10 | 2 | 1 | 1 | 179.5 | 518 | tc422025d | 10 | 3 | 0 | 0 | 219 |
| 465 | 14mc8 | 10 | 1 | 2 | 0 | 183.5 | 492 | 23mbc222o | 10 | 2 | 2 | 0 | 171.5 | 519 | tc521026d | 10 | 3 | 0 | 0 | 188 |
| 466 | 13mc8 | 10 | 1 | 2 | 0 | 181.5 | 493 | 1mbc430n | 10 | 2 | 1 | 1 | 175.8 | 520 | 6mtc3220n | 10 | 3 | 1 | 0 | 189.5 |
| 467 | 12mc8 | 10 | 1 | 2 | 0 | 185.5 | 494 | 26mbc330o | 10 | 2 | 2 | 0 | 160.5 | 521 | 12Cprlmc3 | 10 | 3 | 1 | 1 | 158.3 |
| 468 | 11mc8 | 10 | 1 | 2 | 1 | 170.5 | 495 | 24mbc330o | 10 | 2 | 2 | 0 | 165 | 522 | bc310hxSC5 | 10 | 3 | 0 | 1 | 192.7 |
| 469 | 1mc9 | 10 | 1 | 1 | 0 | 193.6 | 496 | 22mbc321o | 10 | 2 | 2 | 1 | 172.5 | 523 | ds2024d | 10 | 3 | 0 | 2 | 160 |
| 470 | C10 | 10 | 1 | 0 | 0 | 202 | 497 | 1ms44n | 10 | 2 | 1 | 1 | 186.2 | 524 | 38mtc321024o | 10 | 3 | 2 | 0 | 152.5 |
| 471 | bc440d | 10 | 2 | 0 | 0 | 191.5 | 498 | 22mbc222o | 10 | 2 | 2 | 1 | 174.5 | 525 | 334mtc2210h | 10 | 3 | 3 | 2 | 151 |
| 472 | bCpe | 10 | 2 | 0 | 0 | 190 | 499 | 14mbc321o | 10 | 2 | 2 | 1 | 164 | 526 | 133mtc2210h | 10 | 3 | 3 | 2 | 143.5 |
| 473 | bc530d | 10 | 2 | 0 | 0 | 193 | 500 | 13mbc330o | 10 | 2 | 2 | 1 | 160.5 | 527 | 177mtc2210h | 10 | 3 | 3 | 2 | 153 |
| 474 | 3mbc331n | 10 | 2 | 1 | 0 | 182 | 501 | 15mbc321o | 10 | 2 | 2 | 2 | 159.5 | 528 | SCptc3210o | 10 | 4 | 0 | 1 | 174 |
| 475 | 6mbc322n | 10 | 2 | 1 | 0 | 190 | 502 | 225mbc221h | 10 | 2 | 3 | 1 | 162 | 529 | tec52100d | 10 | 4 | 0 | 1 | 155 |
| 476 | 2mbc331n | 10 | 2 | 1 | 0 | 187 | 503 | 1pbc410h | 10 | 2 | 1 | 1 | 172.5 | 530 | pec530000d | 10 | 5 | 0 | 0 | 171 |
| 477 | 3mbc430n | 10 | 2 | 1 | 0 | 178 | 504 | 223mbc221h | 10 | 2 | 3 | 1 | 165.5 | | | | | | | |

[a] cn means the number of carbon atoms; cyc denotes the number of cycles; bn corresponds to the number of branches; and v4 indicates the number of quaternary atoms.

arithmetic step. From the subplots especially *1a2*, it should be correct to declare that size information is one of the structural features hidden in the space spanned by Chi indices. From the former studies, Chi indices generally encode molecule size (as the number of carbon atoms in this data set).

Due to the structural diversity, all decanes are selected to further study the structural features hidden in the descriptors. In the direction *a2* = [0.4997, −0.3255, 0.1629, −0.3670, −0.3249, −0.6144], six indices are projected to be the situation in subplots *1b1* and *1b2*, in which the groups marked by c*0*, c*1*, ..., and c*5* denote the different sets of acyclic, monocyclic, ..., and pentacyclic decanes. The subplots *1b1* and *1b2* are essentially identical, in which the difference is that the *x*-axis of *1b1* and *1b2* is the sequence no. of the listed molecules and the number of cycles of the molecules, respectively. The subplot *1b2* is gotten by using the cycle number, from which the structures with different cycles are distinguished, although the distinctive lines between some groups are not so clear (such as bicyclic and tricyclic structures). In the projection direction, the cyclicity information can be roughly coded by the indices. The same conclusions can be reached from other molecules with different atoms.

Then, all acyclic decanes are chosen to be the data set to search structural features. In the projection direction *a3* = [1.0000, −0.0000, 0.0027, −0.0000, 0.0019, −0.0000], the six Chi indices are projected to show some interesting structure information in subplots *1c1* and *1c2* of Figure 1, in which the lines *b0*, *b1*, ..., and *b5* correspond to the different sets of structures with none, one, ..., and five branches, and another interesting thing is that the structures in the branch groups (such as *b2*, *b3*, ...) are further divided into subgroups (*v40*, *v41*, *v42*, ...), in which the *v40*, *v41*, and *v42* are the set of structures with zero, one, and two quaternary (vertex degree equal to 4) atoms, respectively.

In the next step, the monocyclic decanes are applied in the projections in direction *a4* = [0.8574, −0.1794, −0.2705, −0.0172, 0.3985, 0.0207], which is shown in the subplots *1d1* and *1d2* of Figure 1. The *b0*, *b1*, ..., and *b6*, *v40*, *v41*, and *v42* hold the same meaning to those of the subplots *1c1* and *1c2*.

From the subplots of all acyclic and monocyclic decanes, branching information is, to some extent, described by the six Chi indices, in which a quaternary atom is a special case that two branches are connected on the same atom. In the former studies, Hall and Kier have introduced the difference Chi indices[5,6] as a means of removing the size element so as to emphasize branching information.

What should be pointed out is that the cyclicity is also coded by the multidimensional combination of Chi cluster and path/cluster indices, which was only claimed for the Chi chain indices in the former studies.[5,6]

The structure information such as size, cyclicity, and branching can also be roughly found on other projection directions, and the directions listed in the paper are the directions showing lower "entropy", which means the more the data tends to segregate into clusters. Due to the higher "entropy" of the projection directions, the other structure information such as patterns of adjacency discussed for the individual Chi index is not obvious in this study.

The next interest is which indices and how much an index contributes to a specific feature. Consider the four projection directions of ${}^{0}\chi_{\text{path}}$, ${}^{1}\chi_{\text{path}}$, ${}^{2}\chi_{\text{path}}$, ${}^{3}\chi_{\text{path}}$, ${}^{3}\chi_{\text{cluster}}$, and ${}^{4}\chi_{\text{path}}$:

*a1* = [−0.1869, −0.9613, 0.0052, −0.0274, −0.1994, 0.0222] for size,

*a2* = [0.4997, −0.3255, 0.1629, −0.3670, −0.3249, −0.6144] for cyclicity,

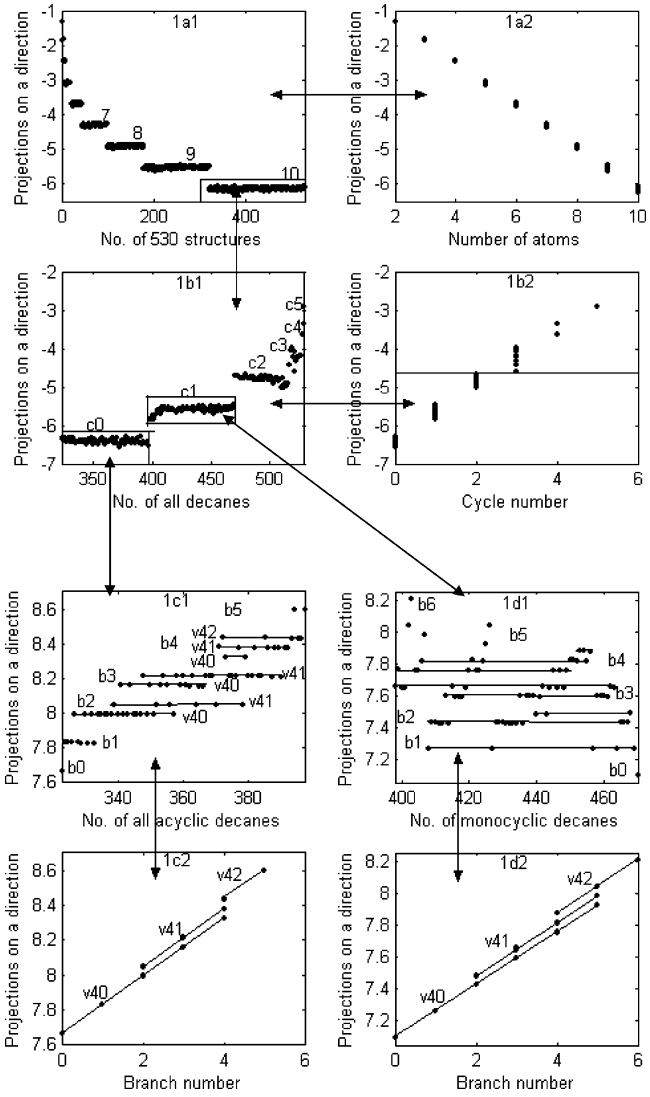


**Figure 1.** Structural features mined out in the space spanned by the six Chi indices (c0, c1, c2; b0, b1, b2; and v40, v41, v42 mean zero, one, two cycles; branches; and quaternary atoms).

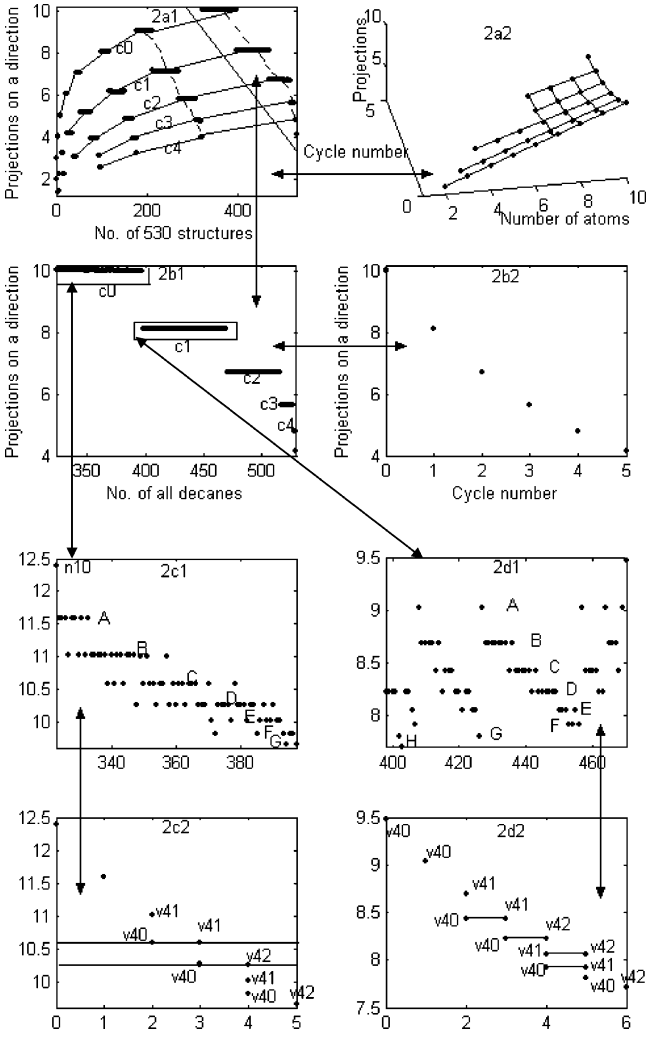


**Figure 2.** Structural features mined out in the space spanned by the four Kappa indices.

*a3* = [1.0000, −0.0000, 0.0027, −0.0000, 0.0019, −0.0000] for branching of acyclic decanes,

and *a4* = [0.8574, −0.1794, −0.2705, −0.0172, 0.3985, 0.0207] for branching of cyclic decanes.

A different index contributes differently to the structure information, in which the branching information is mainly coded by $^{0}\chi_{path}$. From the other directions, almost every index has some contributions to the different structural features. That is, the features are described well by the combination of multidimensional descriptors. In other words, the interpretation problems should involve a multidimensional description of structures.

Another interesting thing is that, from the projection directions, almost every index has different contributions to different structural features. That is, the topological index is commonly a descriptor hiding many finer features into one quantifier, which might be one of the reasons leading to the critique that the structural interpretation of the topological index is difficult.

The projection direction *a3* should be detailed. In that direction, the $^{0}\chi_{path}$ plays a key role, and the other five descriptors contribute a smaller part to the structural feature, in which the role of $^{1}\chi_{path}$ is interesting for it[20] and was claimed to code molecular branching; however, it has no contribution to branching in this projection direction. The contribution of $^{1}\chi_{path}$ on molecular branching might be found on other projection directions, which can be mined out through the rotation of a high-dimensional space in different directions. There is another insight that the different combinations of descriptors can exhibit, to some extent, the same structure features in different directions. The study might provide some helpful results in understanding that the collinearity among topological indices is serious.

**Kappa Index.** In the former study,[7,8] the Kappa shape indices encode intuitively, in combination, significant information on the degree of cyclicity and the degree of centralization/separation in branching. By the same strategy in the above section, the main structural features for different projection directions are discussed below:

In the direction *a1* = [1.0000, −0.0003, 0.0071, −0.0000] for four Kappa indices shows the structure information in subplots *2a1* and *2a2* of Figure 2. In the direction, the information is composed of size and also cyclicity. Only consider line *c0* or *c1* et al.; the information is mainly size. The size information of the Kappa index is different from that of the Chi index, which will mix with the cyclicity information in the Kappa index.

The structural features hidden in all decanes are further studied in the same direction as shown in subplots *2b1* and *2b2* in Figure 2. Consider all decanes in the direction, same as the *a1* direction, *a2* = [1.0000, −0.0003, 0.0071, −0.0000], four Kappa indices are projected to be the situation in subplot *2b2*, from which the cyclicity information can be coded by the indices.

Next, all acyclic decanes are chosen to be the data set to search structural features. In the projection direction *a3* = [0.8548, 0.5120, 0.0007, −0.0851], the four Kappa indices are projected to show some structural information in subplots *2c1* and *2c2* in Figure 2. The line A is for structures with one branch. The line B denotes structures with two branches and one quaternary atom. The line C means structures with both two branches without quaternary atoms and three branches with one quaternary atom. The line D is a set of structures with both four branches with two quaternary atoms and three branches without a quaternary atom. The line E includes structures with four branches with one quaternary atom. The line F corresponds to structures with four branches without quaternary atoms. The line G includes structures with five branches and two quaternary atoms.

Then, the monocyclic decanes are applied in the projections in direction *a4* = [0.6195, −0.7777, −0.0083, 0.1062], which is shown in subplots *2d1* and *2d2* of Figure 2. The line A is for structures with one branch. The line B denotes structures with two branches and one quaternary atom. However, the line C is composed of structures with both three branches with one quaternary atom and two branches without quaternary atom. Similarly, the line D is for structures with both four branches with two quaternary atoms and three branches without a quaternary atom, and the line E is a compilation of structures with both five branches with two quaternary atoms and four branches with one quaternary atom. The line F corresponds to structures with five branches with one quaternary atom and four branches without quaternary atoms. The line G includes structures with both five branches without quaternary atoms. The line H is the structures with six branches and two quaternary atoms.

The above results about Kappa shape indices also support that Kappa indices[7,8] encode, in combination, significant information on the degree of cyclicity and the degree of centralization/separation in branching.

Consider the four projection directions on the four Kappa indices ${}^1\kappa$, ${}^2\kappa$, ${}^3\kappa$, and $\phi$:

*a1* = [1.0000, −0.0003, 0.0071, −0.0000] for size and cyclicity,

*a2* = [1.0000, −0.0003, 0.0071, −0.0000] for size and cyclicity,

*a3* = [0.8548, 0.5120, 0.0007, −0.0851] for branching of acyclic structures,

and *a4* = [0.6195, −0.7777, −0.0083, 0.1062] for branching of cyclic structures.

Similar results to the Chi indices are obtained as follows: (1) a different index contributes differently to the structure information. From the directions *a1*, *a2*, *a3*, and *a4*, almost every index has some contributions (although some indices hold little part) to the different features. That is, the features are described by the combination of multidimensional descriptors. (2) There is also another insight that the different combinations of descriptors can exhibit, to some extent, the same structure features for different directions.

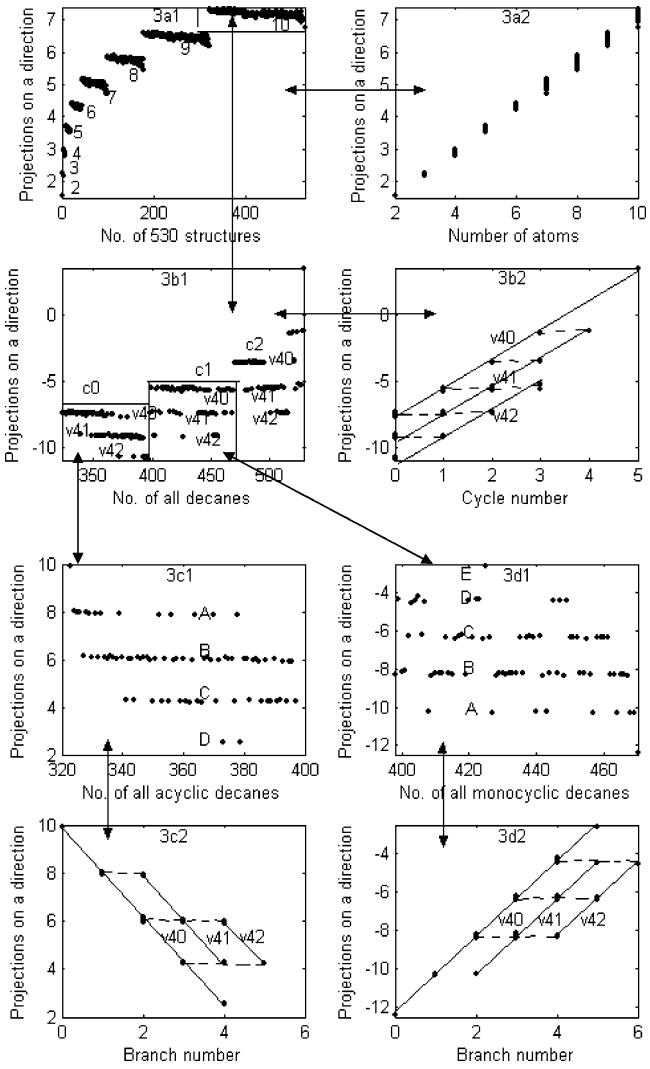


**Figure 3.** Structural features mined out in the space spanned by the four atom-type E-State indices.

From the analysis, the six Chi indices and four kappa indices hold a lot of the same structure information, which might lead to the high collinearity between them.

**Atom-type E-State Index.** There are still a few works paying attention to the structural features coded by atom-type E-State indices. To mine out the structural features hidden in the space spanned by the atom-type E-State indices it might be helpful to understand the indices themselves and study their relationship with other topological indices.

Projection in the direction *a1* = [0.3918, 0.4843, 0.5003, 0.6013] by four atom-type E-State indices shows the size information in subplots *3a1* and *3a2* of Figure 3, in which the integral numbers on the plot is the molecular size. Similar to subplots *1a1* and *1a2* of Chi indices, the difference between the different sizes is changed with an arithmetic step. From the plot, it should also be correct to declare that size information is one of the structural features hidden in the space spanned by atom-type E-State indices. From the definition, the atom-type E-State index is a summation index. Thus, it is expected that size be encoded.

Then, all decanes are selected to further study the structural features hidden in the descriptors. In the direction *a2* = [−0.6638, −0.3783, 0.4583, −0.4541], four atom-type E-State indices are projected to be the situation in subplots

*3b1* and *3b2*, from which the cyclicity information can be roughly coded by the indices. In the different areas, there exist other different groups: $v40$, $v41$, and $v42$. The situation is similar to the quaternary atom subgroup of subplot *1c1* and *1c2* of Chi indices; however, the subgroup of quaternary atoms in this case is based on the cyclicity of molecules, while that of the Chi indices is based on branching of the molecules.

Third, all acyclic decanes are chosen to be the data set to search structural features. In the projection direction *a3* = [0.2453, 0.7699, −0.2569, 0.5302], the four atom-type E-State indices are projected to show some structure information in subplot *3c1* and *3c2* in Figure 3. In the subplot *3c1*, the line A corresponds to both structures with one branch and structures with two branches and one quaternary atom. The line B is a set of structures with two branches without a quaternary atom, with three branches and one quaternary atom, and with four branches and two quaternary atoms. The line C includes structures with three branches without a quaternary atom, with four branches with one quaternary atom, and five branches with two quaternary atoms. The line D is composed of four branches without a quaternary atom.

The main difference between this case and these of both Chi indices and Kappa indices is that the structures with one more quaternary atom will "jump" inversely to the structures with one less quaternary atom. For example, in subplots *3c1* and *3c2* of the atom-type E-State, the line D is composed of four branches without a quaternary atom, with five branches with one quaternary atom, or with six branches with two quaternary atoms, while, in subplot *2c1* and *2c2* of the Kappa indices, the line D is for structures with both four branches with two quaternary atoms and three branches without a quaternary atom.

Fourth, the monocyclic decanes are applied in the projection direction *a4* = [−0.3769, −0.8274, 0.4131, 0.0528], which is shown in subplots *3d1* and *3d2* of Figure 3. Similar to subplot *3c1* and *3c2* but different from subplots *1c1*, *1c2*, *2c1*, and *2c2*, the line A corresponds to both structures with one branch and structures with two branches and one quaternary atom. The line B is a set of structures with two branches without a quaternary atom, with three branches and one quaternary atom, and with four branches and two quaternary atoms. The line C includes structures with three branches without a quaternary atom, with four branches with one quaternary atom, and five branches with two quaternary atoms. The line D is composed of four branches without a quaternary atom, with five branches with one quaternary atom, or with six branches with two quaternary atoms.

Consider the four projection directions of the four atom-type E-State indices:

*a1* = [0.3918, 0.4843, 0.5003, 0.6013] for size,

*a2* = [−0.6638, −0.3783, 0.4583, −0.4541] for cyclicity and branching,

*a3* = [0.2453, 0.7699, −0.2569, 0.5302] for branching of acyclic decanes,

and *a4* = [−0.3769, −0.8274, 0.4131, 0.0528] for branching of cyclic decanes.

The different index contributes differently to the structure information. From the directions *a1*, *a2*, *a3*, and *a4*, almost every index has some contributions to the different structural features. That is, the structural features are described by the combination of multidimensional descriptors.

From the above analysis, the structural features coded by the atom-type E-State indices are mainly size, cyclicity, branching, and the number of quaternary atoms, which are found out by the rotation of high dimension space in different directions.

**Applications of Interpretation Information.** Similar to the prior study,[1] we tried to use the structural features to discover how the boiling points change with the variation of structural features by the statistics way. It is well-known that with the addition of the number of atoms in a structure, the boiling points will, in general, increase. It can also be roughly predicted that with the increase of the branch number in a structure, the boiling points will, in general, decrease. Next, we will consider the effects of the cycle number and the number of quaternary atoms.

To simplify the case, the size effects are not taken into account, that is, only the decanes are selected in the further analysis. At first, the different boiling points of decanes with the same cycle number are classified, and the boiling points of them are added together to get a value. Then, the value is divided to obtain the average boiling point for a specific cycle number. The changes in the different cycles show that with the addition of a cycle number in a structure, the boiling points will, in general, increase compared with the acyclic structures. However, the average boiling points of the multicyclic structures are not statistically increasing or decreasing, which might be caused by the incomplete sample structures.

The structural feature such as the number of quaternary atoms ($v4$) is mined out in the present work and to study the effects of $v4$ should be essential. Due to the structure diversity or complexity of the size, cyclicity, branch number, branching position, et al., it is difficult to select a case to study the effects of $v4$. However, we have tried to apply a new way to test whether $v4$ holds some influence on the modeling. The data set is the bicyclic decanes with one branch, in which there are two cases (**X** with none $v4$ and **Y** with one $v4$). Then the six Chi indices are used to model the boiling points of **X** and Y. For the **X** set, the regression results are $R = 0.9938$, $s = 3.0523$, $F = 199.857$ and the maximum absolute residual (MAR) = 5.84, and those for the **Y** set are $R = 0.9977$, $s = 2.9246$, $F = 398.3681$, and MAR = 5.5. When both sets are put together to regress, the results are $R = 0.9940$, $s = 4.0164$, $F = 452.8604$, and especially MAR = 10.32. The same results are also obtained in other cases. The phenomenon shows that the role of $v4$ is relatively important, which should be partitioned further in some situations.

Briefly, by projection pursuit combining with the number theory method, the structural features hidden in the space spanned by several topological indices are mined out, and their effects are applied in discovering chemical knowledge, which provides some new thoughts about the nature of the topological index by using a multidimensional idea.

## CONCLUSION

The structural features within the three series of indices hold some of the same and different aspects. The main features coded by them are size, cyclicity, and branching (including a quaternary atom). First, the structural features are mined out by the combination of several indices. Second,

although the three series of indices are defined differently, the structural features described by them are mainly the same, which indicate that the collinearity might be a serious problem among the applications of the topological indices. And then, to study the relationship between the different series of indices is valuable, which will appear in our subsequent paper.

## ACKNOWLEDGMENT

This project is financially supported by National Natural Scientific Foundation Committee (NSFC) of P. R. China (No. 20235020 and 20175036). The valuable comments and suggestions for improving the paper from the referee are highly appreciated.

## REFERENCES AND NOTES

(1) Hu, Q. N.; Liang, Y. Z.; Peng, X. L.; Yin, H.; Fang, K. T. Structural Interpretation of Topological Index. 1. External Factor Variable Connectivity Index (EFVCI). *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 437−446.

(2) Kartrizky, A. R.; Chen, K.; Maran, U.; Carlson, D. A. QSPR correlation and predictions of GC retention indexes for methyl-branched hydrocarbons produced by insects. *Anal. Chem.* **2000**, *72*, 101−109.

(3) Randić, M.; Zupan, J. On Interpretation of well-known Topological Indices. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 550−560.

(4) Randić, M.; Balaban, A. T.; Basak, S. C. On structural interpretation of several distance related topological indices. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 593−601.

(5) Kier, L. B.; Hall, L. H. *Molecular connectivity in chemistry and drug research*; Academic Press Inc.: New York, 1976.

(6) Kier, L. B.; Hall, L. H. *Molecular connectivity in structure−activity analysis*; Research Studies Press Ltd.: New York, 1986.

(7) (a) Kier, L. B. A shape index from chemical graphs. *Quantum Struct. Act. Relat.* **1985**, *4*, 109−116. (b) Kier, L. B. Shape indexes of orders one and three from molecular graphs. *Quant. Struct. Act. Relat.* **1986**, *5*, 1−7.

(8) (a) Kier, L. B. Kappa shape indices for similarity analysis. *Med. Chem. Res.* **1997**, *7*, 8−12. (b) Kier, L. B.; Hall, L. H. The kappa indices for modeling molecular shape and flexibility. In *Topological indices and related descriptors in QSAR and QSPR*; Devillers, J., Balaban, A. T., Eds.; Gordon and Breach Science Publishers: 1999; pp 455−490.

(9) (a) Kier, L. B.; Hall, L. H. An electrotopological state index for atoms in molecules. *Pharm. Res.* **1990**, *7*, 801−807. (b) Hall, L. H.; Kier, L. B. *Molecular structure description: the electrotopological state*; Academic Press: 1999. (c) Hall, L. H.; Kier, L. B.; Brown, B. B. Molecular similarity based on novel atom-type E-State indices. *J. Chem. Inf. Comput. Sci.* **1995**, *35*, 1074−1080. (d) Kellogg, G. E.; Kier, L. B.; Gaillard, P.; Hall, L. H. E-State fields: Application to 3-D QSAR. *J. Comput.-Aided. Mol. Des.* **1996**, *10*, 513−520. (e) Hall, L. H.; Vaughn, A. T. QSAR of phenol toxicity using E-State and Kappa shape indices. *Med. Chem. Res.* **1997**, *7*, 407−416. (f) Hall, L. H.; Story, C. T. Boling point and critical temperature of a heterogeneous data set: QSAR with atom-type E-State indices using artificial neural networks. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1004−1014. (g) Hall, L. H.; Kier, L. B. *The E-State as the Basis for Molecular Structure Space Definition and Structure Similarity* **2000**, *40*, 784−791.

(10) (a) Tetko, I. V.; Tanchuk, V. Y.; Kasheva, T. N.; Villa, A. E. Estimation of aqueous solubility of chemical compounds using E-state indices. *J. Chem. Inf. Comput. Sci.* **2001**, *41*, 1488−1493. (b) Huuskonen, J. J.; Livingstone, D. J.; Tetko, I. V. Neural network modeling for estimation of partition coefficient based on atom-type electrotopological state indices. *J. Chem. Inf. Comput. Sci.* **2000**, *40*, 947−55.

(11) (a) Friedman, J. H.; Tukey, J. W. A projection pursuit algorithm for exploratory data analysis. *IEEE Trans. Comput.* **1974**, *C-23*, 881−889. (b) Friedman, J. H. Exploratory projection pursuit. *J. Am. Stat. Assoc.* **1987**, *82*, 249−266.

(12) Huber, P. J. Projection Pursuit (with discussion). *Annals Statistics* **1985**, *13*, 435−475.

(13) Du, Y. P.; Liang, Y. Z.; Yun, D. Data Mining for Seeking an Accurate Quantitative Relationship between Molecular Structure and GC Retention Indices of Alkenes by Projection Pursuit. *J. Chem. Inf. Comput. Sci.* **2002**, *42*, 1283−1292.

(14) Tashiro, Y. On methods for generating uniform points on the surface of a sphere. *Annals Institute Statistical Math.* **1977**, *29*, 295−300.

(15) Fang, K. T.; Wang, Y.; Wong, K. L. *A new method of generating an NT-net on the unit sphere*; Technical Report MATH-002; Hong Kong Baptist College: 1992.

(16) Niederreiter, H. Pseudo-random numbers and optimal coefficients. *Adv. Math.* **1977**, *26*, 99−181.

(17) Hua, L. K.; Wang, Y. *Applications of number theory to numerical analysis*; Spring-Verlag and Science Press: Berlin and Beijing, 1981.

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

(19) Hu, Q. N.; Liang, Y. Z.; Wang, Y. L.; Guo, F. Q.; Huang, L. F. The basic principles of heuristic queue notation and its applications in calculating matrixes and topological index for topological graphs. *Comput. Appl. Chem. (in Chinese)* **2003**, *20*, 386−390.

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

CI049973Z