Journal of Computer-Aided Molecular Design (2005) 19: 651–660
DOI 10.1007/s10822-005-9010-6

Can topological indices transmit information on properties but not on structures?

Alexandru T. Balaban[a,*]
[a]*Texas A&M University at Galveston, 5007 Avenue U, Galveston, TX 77551, USA*

Received 11 May 2005; accepted 28 July 2005

Key words: Molecular descriptor degeneracy, three topological index generations, QSAR/QSPR, reverse engineering, inverse problem

Summary

A survey of several topological indices (TIs) is provided according to the nature (integers or real numbers) of the local vertex invariants (LOVIs) and of the resulting molecular descriptor (TI). The 1st generation TIs such as the Wiener index when both the LOVIs and the TI are integers have a very high degeneracy. This fact can become an asset for the problem under discussion: when confronted with the "inverse problem" (reverse engineering) such indices lead to a combinatorial explosion of possible solutions. On the other extreme, TIs with very low degeneracy may also offer the possibility to transmit information on properties but not on structures, because they may be too difficult to lend themselves to reverse engineering in a reasonable amount of time. Several such indices are discussed: the novel second-generation index G (derived from the average distance-based connectivity index J in order to include a dependence on graph size and cyclicity), and third-generation indices such as the triplet TI denoted by $DN^2S(4)$ or the Kier–Hall index TOTOP. The intercorrelation of these indices is discussed: G is almost linearly-correlated with $DN^2S(4)$, and shows a different type of correlation with TOTOP.

Introduction

Topological indices (TIs) are numbers associated with chemical structures for the purpose of QSPR/QSAR studies. Their use in lead discovery and optimization for drug design and in correlating and predicting physical properties has been documented in books [1–11] and reviews [12–16]. Three generations of TIs can be delineated in terms of the nature (integer or real numbers) of the local (atom or bond) invariants (LOVIs) and the resulting global (molecular) TI, as seen in Table 1.

*To whom correspondence should be addressed. Tel: 409-741-4313; Fax: 409-740-4787; E-mail: balabana@tamug.edu
Presented in part at the 229th ACS National Meeting San Diego, March 13, 2005

If two or more chemical graphs have the same TI, then the TI is said to be degenerate. So far no TI characterizes a graph up to isomorphism. The degeneracy of TIs decreases, and the discrimination power increases, from the first to the second and third generations.

Brief review of several topological indices

A few TIs will be mentioned to serve as examples for the three generations of TIs. We start with 1st generation TIs, which show a high degeneracy.

(1) *The Wiener index W* (the first TI) is based on the topological distance matrix [17]:

Table 1. Three generations of topological indices.

Generation	Local vertex invariants (LOVIs)	TIs
First generation (e.g. Wiener, Hosoya, centric index)	Integers	Integers
Second generation (e.g. Randic, Kier–Hall, Balaban's *J*)	Integers	Real Nos.
Third generation (e.g. triplet indices, BCUT indices)	Real Nos.	Real Nos.

$$W = 1/2\sum_i S_i = 1/2\sum_i\sum_j d_{ij}$$

where d_{ij} denotes an entry in the distance matrix, and S_i denotes the *distance sum* for the *i*-th row or column corresponding to vertex *i*. Topological distances are the number of edges on the shortest path between two vertices. Taking the three pentane isomers as examples, *W* values are 20 for *n*-pentane, 18 for isopentane (2-methylbutane), and 16 for neopentane (2,2-dimethylpropane).

(2) *Centric indices* for acyclic graphs (trees) are based on the number of vertices of degree 1 (endpoints) in hydrogen-depleted graphs that are deleted (pruned) at each stage till one reaches the center (a single vertex) or the bicenter (a pair of adjacent vertices) [18]. The simplest centric index *B* is the sum of squares in the pruning sequence. Figure 1 illustrates the centric indices of the three isomers of pentane: *B* is 9 for *n*-pentane, 13 for isopentane (2-methylbutane), and 17 for neopentane (2,2-dimethylpropane), showing an inverse variation with branching when compared with *W*. Centers of polycyclic graphs have more complicated definitions [19].

(3) *The Hosoya index Z* is obtained by summing the number of ways in which *k* edges may be chosen so that no two of them are adjacent [20]. For trees there exists a simple relationship between *Z* and the values of the coefficients in the characteristic polynomial of the graph. As examples, *Z* values are 8 for *n*-pentane, 7 for isopentane (2-methylbutane), and 5 for neopentane (2,2-dimethylpropane).

(4) Discussing now 2nd generation TIs, *Randić's connectivity index* χ is based on the product of vertex degrees v_i of edge endpoints for each edge [21]:

$$\chi = \sum_{i-k}(v_i v_k)^{-1/2}$$

where the summation extends over all edges *i-k*. Because the vertex degrees in organic hydrogen-depleted molecules are integers 1, 2, 3, or 4, there are only 10 distinct combinations for the products of degree endpoints. The alkanes with >7 carbons have degenerate χ indices. As examples, χ values are 2.4142 for *n*-pentane, 2.2701 for isopentane (2-methylbutane), and 2.0000 for neopentane (2,2-dimethylpropane), showing the same decreasing trend with increasing branching as index *B*. More recently, Randić has introduced variable connectivity indices where the main diagonal of the adjacency matrix contains a column vector with variables for various types of heteroatoms or even carbon atoms (primary, secondary, etc.), and the experimental data are used to obtain the best correlation factor [22].

(5) The *Kier–Hall topological indices* [6–8] ($^k\chi$) are generalizations of Randić's connectivity index obtained by paths of length *k* (an edge is a path of length 1). For molecules with heteroatoms, valence-chi indices $^k\chi^v$ have been

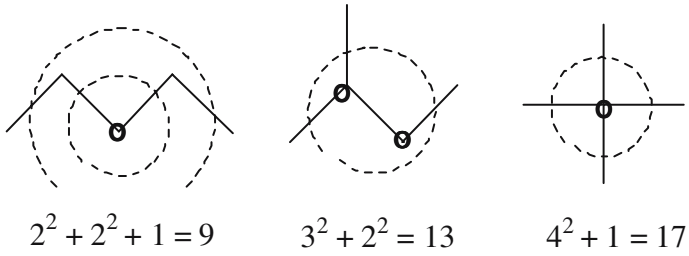

Figure 1. The pruning process for obtaining centric indices of the three pentane isomers.

devised by assigning fixed weights to vertices. More recently, Kier and Hall have introduced electro-topological state indices [8].

(6) *Several information-theoretical indices* have been devised by Bonchev and Trinajstić (I_D^W, I_D^E) [23] and by Basak and coworkers (IC, CIC, SIC) [24–27]. Then other information-theoretical indices (U, V, X, Y) were suggested by Balaban and Balaban [28].

(7) *The average distance connectivity index* (J) (Balaban's index) was designed so as not to increase rapidly with the graph size, like all other preceding TIs [29, 30]. It used a similar summation for all edges *i-k* as Randić's connectivity index χ, but the LOVIs were the distance sums S_i, integers, which are less degenerate and unlimited in magnitude; in addition, the index J is normalized for graph size and cyclicity by means of the number E of edges and of the cyclomatic number ($R = E-N + 1$):

$$J = [E/(R+1)]\sum\nolimits_{i-k}(S_iS_k)^{-1/2}$$

In Figure 2, the three pentane isomers with vertex degree (v_i) and distance sum (S_i) values are shown. It is evident that v_i as LOVIs are much more degenerate than S_i: there are two types of secondary carbon atoms in *n*-pentane and two types of primary carbon atoms in isopentane, yet v_i values do not discriminate them, but S_i values do.

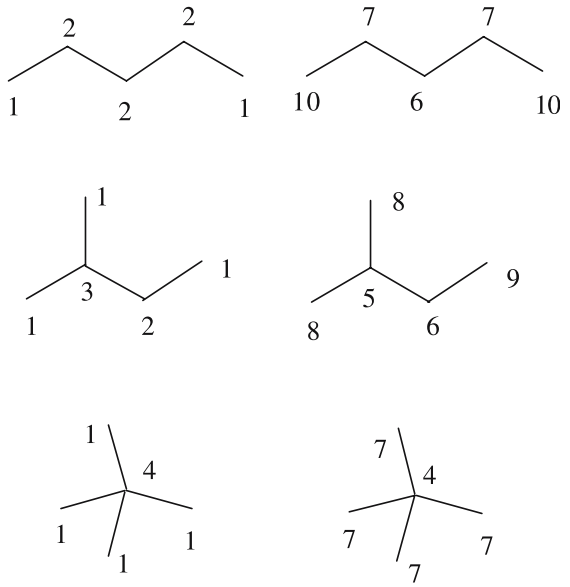

Figure 2. Two types of LOVIs for the three pentane isomers, *n*-pentane (top), isopentane (middle) and neopentane (bottom): vertex degrees (v_i) on the left and distance sums (S_i) on the right.

As a consequence of this normalization, for an infinite linear chain (polyethylene) $J = \pi$, and for infinite polypropylene $J = 3\pi/2$ [31]. For alkanes, J starts to be degenerate when there are 12 carbon atoms, and for cyclic graphs with 8 carbon atoms [32]. Such graph pairs are presented in Table 2 and Figure 3.

It is easy to compute index J for compounds containing heteroatoms and/or multiple bonds: the LOVI for the corresponding vertex is multiplied by a parameter that characterizes either the electronegativity or the covalent radius of the heteroatom relatively to carbon. Then, the result replaces the distance sum for that vertex. Bond parameters are computed by weighting schemes in terms of bond multiplicity and of either the relative electronegativity or the relative covalent radii of the two atoms involved in the bond [33, 34]. An interesting observation is that index J orders alkanes like index W, [35, 36] and also similarly to Bertz's line-graph-derived descriptor [37].

It was rigorously proved that:

Table 2. The 4 graphs with 8 vertices (numbered 1–4) and the 12 graphs with 12 vertices (numbered 5–16) that have pairwise degenerate values for index J. Two other TIs, namely $DN^2S(4)$ and TOTOP, are also indicated, and they will be discussed later.

Graph	W	J	$DN^2S(4)$	TOTOP
1	60	2.422	48.402	27.548
2	60	2.422	48.402	27.496
3	56	2.161	56.808	35.225
4	56	2.161	56.794	35.158
5	216	3.752	62.207	34.880
6	216	3.752	62.174	34.803
7	196	4.253	69.213	35.780
8	196	4.253	69.222	35.832
9	200	4.135	67.512	35.574
10	200	4.135	67.491	35.506
11	225	3.575	59.763	34.145
12	225	3.575	59.746	34.131
13	216	3.773	62.610	34.384
14	216	3.773	62.603	34.3843
15	207	3.954	64.943	34.955
16	207	3.954	64.928	34.940

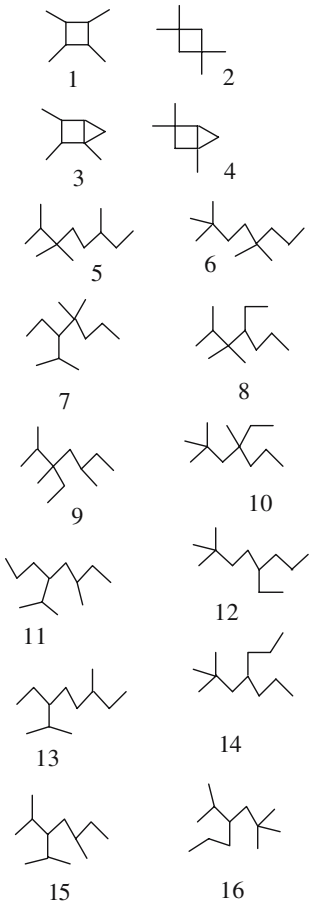

Figure 3. Graphs with 8 and 12 vertices with pairwise degenerate J values.

For almost all connected graphs G there is a non-isomorphic graph G' such that $J(G) = J(G')$ [38].

If such a TI with low degeneracy has this property, then *a fortiori* the other 2nd generation TIs described here before it have the same property.

(8) *Assembling LOVIs into a Triplet TI*. Triplet TIs are obtained by converting a symmetrical matrix (e.g. **A** or **D**) into a system of linear equations whose solutions are real number vertex invariants (LOcal Vertex Invariants, LOVIs). This is done by combining the matrix with two column vectors (main diagonal and free term) adding chemical (e.g. atomic number Z), topological (e.g. vertex degree V or distance sum S), or numerical information (e.g. the number N of atoms in the hydrogen-depleted graph, or its square N^2) [39–42].Then the LOVIs are assembled into a TI by one of the operations below:

1. Summation, $\sum_i x_i$;
2. Summation of squares, $\sum_i x_{i^2}$;
3. Summation of reciprocal square roots, $\sum_i x_i^{-1/2}$;
4. Sum of inverse square root of cross-product over edges ij, $\sum_{ij}(x_i x_j)^{-1/2}$;
5. Averaged product, $N[\Pi_i x_i]^{1/N}$

An identity graph has no two structurally equivalent vertices. It can be shown that the smallest identity tree is unique and has seven vertices. This identity tree (Figure 4) is used to

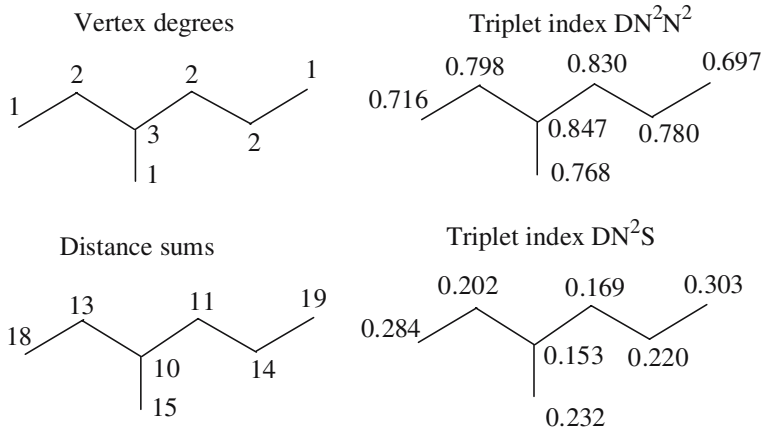

Figure 4. Correlation between LOVIs (for the smallest identity tree) for triplets DN^2S and S. The third-generation triplet TI $DN^2S(4)$ is based on real-number LOVIs DN^2S, and the second-generation TI G is based on integer distance sums S. The smallest identity tree with two of its integer LOVIs (vertex degrees, v_i, and distance sums, S_i on the left-hand side), as well as with two real-number LOVIs derived from triplet indices (on the right-hand side).

$$
\begin{aligned}
49x_1 + x_2 + 2x_3 + 3x_4 + 4x_5 + 5x_6 + 3x_7 &= 18\\
x_1 + 49x_2 + x_3 + 2x_4 + 3x_5 + 4x_6 + 2x_7 &= 13\\
2x_1 + x_2 + 49x_3 + x_4 + 2x_5 + 3x_6 + x_7 &= 10\\
3x_1 + 2x_2 + x_3 + 49x_4 + x_5 + 2x_6 + 2x_7 &= 11\\
4x_1 + 3x_2 + 2x_3 + x_4 + 49x_5 + x_6 + 3x_7 &= 14\\
5x_1 + 4x_2 + 3x_3 + 2x_4 + x_5 + 49x_6 + 4x_7 &= 19\\
3x_1 + 2x_2 + x_3 + 2x_4 + 3x_5 + 4x_6 + 49x_7 &= 15
\end{aligned}
$$

$$
\begin{aligned}
49x_1 + x_2 + 2x_3 + 3x_4 + 4x_5 + 5x_6 + 3x_7 &= 49\\
x_1 + 49x_2 + x_3 + 2x_4 + 3x_5 + 4x_6 + 2x_7 &= 49\\
2x_1 + x_2 + 49x_3 + x_4 + 2x_5 + 3x_6 + x_7 &= 49\\
3x_1 + 2x_2 + x_3 + 49x_4 + x_5 + 2x_6 + 2x_7 &= 49\\
4x_1 + 3x_2 + 2x_3 + x_4 + 49x_5 + x_6 + 3x_7 &= 49\\
5x_1 + 4x_2 + 3x_3 + 2x_4 + x_5 + 49x_6 + 4x_7 &= 49\\
3x_1 + 2x_2 + x_3 + 2x_4 + 3x_5 + 4x_6 + 49x_7 &= 49
\end{aligned}
$$

Figure 5. Conversion of the distance matrix **D** into a system of linear equations by inserting a unit vector N^2 on the main diagonal and a unit vector as the free term. This latter vector is either the distance sum S_i for obtaining triplet LOVIs DN^2S (upper system of linear equations), or triplet LOVIs DN^2N^2 (lower system). The LOVIs are the solutions x_i of these equation systems, as displayed in Figure 4.

exemplify the fact that integer-number LOVIs vertex degrees are degenerate, but distance sums S are all distinct. The linear equation systems shown in Figure 5 based on the distance matrix afford triplet LOVIs DN^2S with $N^2 = 7^2 = 49$ on the main diagonal and distance sums S as the free terms. If the same equations have also $N^2 = 49$ as the free terms, then one obtains DN^2N^2 LOVIs (Figure 5). Real-number LOVIs can then be converted into triplet TIs via operation (4) affording for instance the index $DN^2S(4)$.

The real-number LOVIs obtained with the triplet index DN^2N^2 vary quasi-similarly with the highly degenerate integer vertex degrees, whereas the real-number LOVIs obtained with the triplet index DN^2S correlate linearly with the integer distance sums S having a high correlation factor. In other words, the triplet operation DN^2S converts integer-number LOVIs (S) into real-number LOVIs which have a much lower chance of being degenerate.

A new topological index (G) related to J

A topological index G derived from J that does increase with graph size is:

$$
\begin{aligned}
G &= N^2J(R+1)/(N+R+1)\\
&= N^2J(E-N+2)/(E+2)\\
&= [EN^2/(N+R+1)]\sum\nolimits_{i-k}(s_is_k)^{-1/2}
\end{aligned}
$$

The extra terms of G alongside with J make G dependent not only on branching (like J), but also on graph size and cyclicity. Being derived from J by operations involving only the numbers N of vertices and E of edges, and the cyclomatic number R, G has the same degeneracy as J. More details about index G are to be found in a separate paper [43].

Triplet and Kier–Hall indices with very low degeneracy correlated with index G

There is a need for a molecular descriptor that shares the attributes of the Wiener index without its high degeneracy. We looked for a *triplet TI* that has these properties and found out that $DN^2S(4)$ is such a molecular descriptor. It is based on real-number LOVIs (DN^2S) that correlate fairly well with distance sums S_i. As a consequence, the third-generation index $DN^2S(4)$ with an extremely low degeneracy presents a high correlation with the index G that, being derived from index J, is a second-generation TI. The new index $G = N^2J(R+1)/(N+R-1)$ is almost linearly correlated with the triplet index $DN^2S(4)$ because, the LOVIs on which these two TIs are based, namely the distance sums S_i (integers) and the DN^2S LOVIs (real numbers), are almost linearly correlated. The high value of the correlation factor (Figure 6)

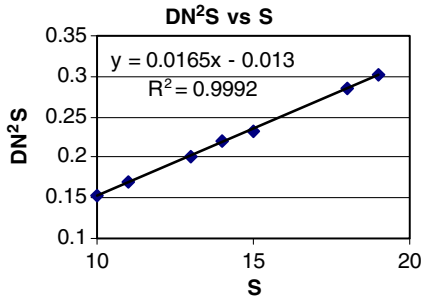

Figure 6. Correlation between LOVIs (for the smallest identity tree) S_i and DN^2S for triplets. The third-generation triplet TI $DN^2S(4)$ is based on real-number LOVIs DN^2S, and the second-generation TI G is based on integer distance sums S.

proves this assertion. Both J and $DN^2S(4)$ are based on the same operation (4).

From Figure 6 one can see that the integer LOVIs S_i show an almost perfect linear correlation with the real number LOVIs obtained from the triplet system of linear equations denoted by DN^2S. One can state that this ''triplet operation'' DN^2S converts integer-number LOVIs into almost linearly-correlated real-number LOVIs. However, one may see from Figure 7 that other triplet LOVIs such as DN^2N^2 evidence an inverse (again almost linear) correlation with the same integer LOVIs S_i for the same identity tree.

Figure 8 confirms that indeed there is a good linear correlation between TIs G and $DN^2S(4)$, by displaying the straight line for the plot that includes the 92 cyclic and acyclic constitutional graphs of alkanes and cycloalkanes with C_4–C_6 chains and the 62 constitutional graphs of alkanes with C_7–C_9 chains.

There are also other 3rd generation TIs that have low degeneracy and can be compared with the triplet index $DN^2S(4)$, such as the Kier–Hall index TOTOP (TOtal TOPological index) [44]. This index is one of the indices included in the Molconn-Z program [45]. One can observe from Figure 9 that for each isomer series of alkanes C_4–C_9, there is an approximately linear correlation between TOTOP and $DN^2S(4)$, indicating that with increasing size and branching of the hydrogen-depleted graphs both these TIs increase.

When in addition to the 72 constitutional isomers of alkanes C_4–C_9 one considers also the 82 cyclic and polycyclic graphs C_4–C_6, one sees in Figure 10 that differences in cyclicity can also be observed: now all alkanes appear in the left-hand lower portion of the plot, whereas with increasing number of rings in the polycyclic graphs the upper right-hand portion shows a steep non-linear increase in the TOTOP values. Both TIs have large values for systems with large numbers of rings.

In addition to the indices discussed so far, the BCUT indices also have been reported by Burden and by Pearlman to possess very low degeneracy, but details should be obtained from the bibliography [46–48].

Reverse engineering (the ''inverse problem'') of TIs (finding chemical structures from one TI). Combinatorial explosion in inverse problems

Several groups of authors have approached the ''inverse problem'' of finding chemical structures corresponding to given values or ranges of particular TIs. The first is the Moscow school around Academician Nikolai Serafimovich Zefirov, that includes Drs Palyulin, Skvortsova, and Baskin [49–60]. Another remarkable group is around Dr Jean-Loup Faulon [61–68]. Professors Lemont B. Kier and Lowell H. Hall have published several papers on this topic [69–71]. A fourth group is constituted by research undertaken by Drs Vladimir Kvasnička and Jiří Pospíchal [72–75]. Finally, other authors have published one paper each in this area [76–79].

This is the crux of the matter in the present context: whether it is possible to exclude the possibility of this ''reverse engineering'' (or to make it as difficult as possible). One wishes to be

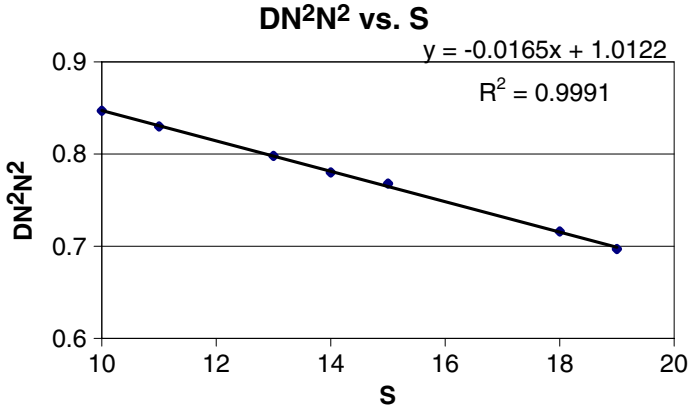

Figure 7. Correlation between the integer LOVIs S_i and the real-number LOVIs DN^2N^2 for the smallest identity tree.

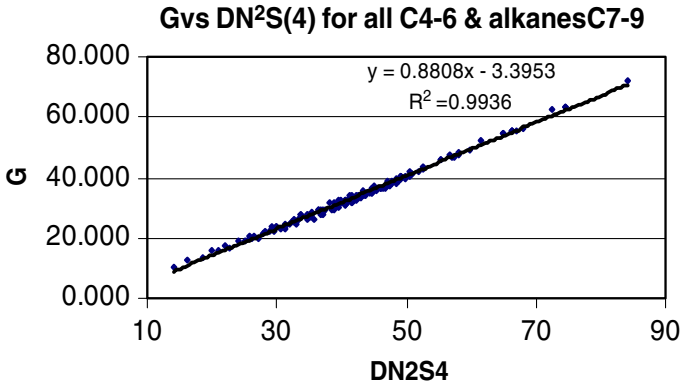

Figure 8. For 154 hydrocarbons (alkanes C_4–C_9 and cyclic structures C_4–C_6) there is a very good correlation between the 2nd generation TI *G* and the 3rd generation TI $DN^2S(4)$.

able to convey information on some properties of chemical compounds without allowing their structures to be guessed. This entails two contradictory requirements: to use molecular descriptors that are useful for carrying information on the given property, but are not amenable to be "cracked" for structural information by algorithms that take a finite amount of time.

There appear to be two possibilities of using TIs for this purpose that may be explored:

(i) Using a TI with such a high degeneracy that the inverse problem becomes intractable due to the "combinatorial explosion" of structures compatible to the value or a range of values for the given TI. One example is provided by first-generation TIs that can be cracked by algorithms described by the above authors, but owing to the high degeneracy, so many structures can be found that for chemical companies "there may be safety in numbers". One candidate among the TIs that are often used in QSAR studies is the Wiener index, and sometimes it is used either alone or with a derived descriptor [80–85].

(ii) Using the other extreme, a TI with such a low degeneracy that a structure compatible with its value would be very hard to detect, especially if this value results from convolute operations or from encoding of information relative to heteroatoms, bond multiplicity, or other structural details. To guarantee that such a descriptor would also be useful for structure-

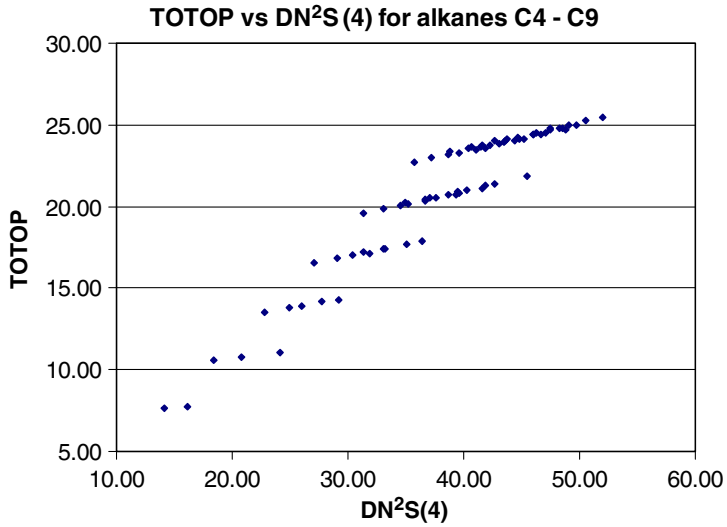

Figure 9. For all alkanes C_4–C_9, the Kier–Hall 3rd generation index TOTOP presents for each number of carbon atoms separate linear correlations with the 3rd generation triplet index $DN^2S(4)$.

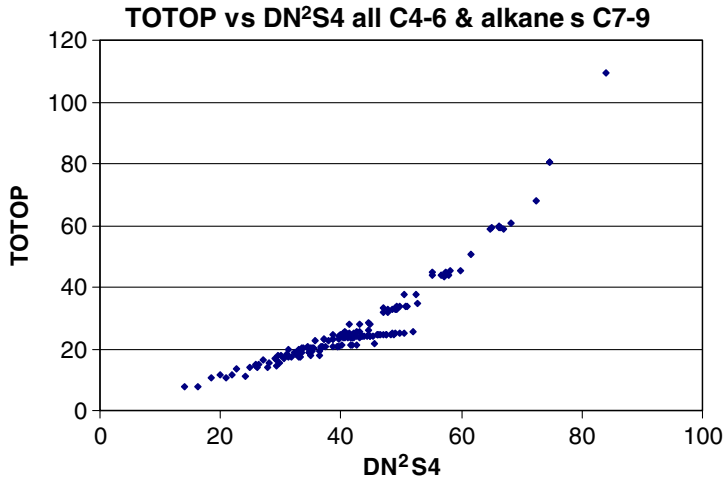

Figure 10. When cyclic systems with C_4–C_6 are added to the preceding set of hydrocarbons, one sees that the correlation between TOTOP and $DN^2S(4)$ TIs becomes non-linear.

property correlations, it would be useful to know that it varies like a descriptor that has higher degeneracy, such as *G*. Third-generation topological indices such as $DN^2S(4)$, TOTOP, or BCUT may have a chance for fulfilling this requirement, i.e. for communicating data on physical–chemical properties or biological activities, without worrying about disclosing the chemical structures. By changing the main-diagonal vector in $DN^2S(4)$ from N^2 to another set of sufficiently large (possibly encoded) numbers, a wide variety of other possible triplet TIs may be invented for even better protection of the confidential information.

Conclusions

The aim of the present discussion was:

(1) to discuss molecular descriptors (topological indices, TIs) that are as far as possible uniquely associated with chemical structures, i.e. TIs with low degeneracy;
(2) these TIs should have good correlational ability, so that they might be used to convey information on chemical, physical, or biological properties;
(3) these TIs should, however, lead to inverse problems that would be extremely hard to solve in a reasonable time;
(4) several such TIs exist, and many more could be designed.

The exchange of information will be stimulated as a consequence of the recently announced effort by the National Institutes of Health to pay for and carry out clinical trials of experimental drugs in order to make it easier for academia and pharmaceutical companies to cooperate in bringing new drugs to the market, starting with treatments for schizophrenia.

In conclusion, there are some certainties (no analytical solutions or specific algorithms exist to reverse-engineer molecular structures from *most* of the 2nd-generation and probably *all* 3rd-generation TIs), as well as some conjectures or educated guesses that have to be tested in practice. One such conjecture is that a single convoluted 3rd-generation TI, which will be hard to reverse-engineer into structure(s) in a reasonable time, may be used for communicating information on properties (as long as the molecular formula, molecular weight and the nature of heteroatoms are kept secret). However, it may be possible that stochastic techniques (such as Monte Carlo, Genetic Algorithms, and Evolutionary Computing) could still be used by hackers with some chance of success. Thus, a new challenge faces specialists in inventing topological indices – to devise molecular descriptors that are hard to crack either by deterministic algorithms or by stochastic techniques.

References

1. Devillers, J. and Balaban, A.T. (Eds.), Topological Indices and Related Descriptors in QSAR and QSPR. Gordon & Breach, Amsterdam, 1999.
2. Todeschini, R. and Consonni, V. Handbook of Molecular Descriptors. Wiley–VCH, New York, 2000.
3. Karelson, M. Molecular Descriptors in QSAR/QSPR. Wiley–Interscience, New York, 2000.
4. Bonchev, D. Information Theoretic Indices for Characterization of Chemical Structure. Research Studies Press–Wiley, Chichester, 1983.
5. Trinajstić, N. Chemical Graph Theory 2nd ed. CRC Press, Boca Raton, 1992.
6. Kier, L.B. and Hall, L.H. Molecular Connectivity in Chemistry and Drug Research. Academic Press, New York, 1976.
7. Kier, L.B. and Hall, L.H. Molecular Connectivity in Structure-Activity Analysis. Research Studies Press, Letchworth, 1986.
8. Kier, L.B. and Hall, L.H. Molecular Structure Description: The Electrotopological State. Academic Press, San Diego, 1999.
9. Randić, M., In Encyclopedia of Computational Chemistry (P. v. R. Schleyer et al., Eds.), Wiley, Chichester, 1998, pp. 3018–3032.
10. Balaban, A.T., In Encyclopedia of Analytical Chemistry (R.A. Meyers, Ed.), Wiley, Chichester, 2000, Vol. 8, pp. 7288–7311.
11. Diudea, M.V. QSPR/QSAR Studies by Molecular Descriptors. Nova Science Publ, Huntington, NY, 2001.
12. Pogliani, L., Chem. Rev., 100 (2000) 3827.
13. Estrada, E. and Uriarte, E., Curr. Med. Chem., 8 (2001) 1573.
14. Gozalbes, R., Doucet, J.P. and Derouin, F., Curr. Drug. Targets–Infectious Disorders, 2 (2002) 93.
15. Balaban, A.T., Motoc, I., Bonchev, D., and Mekenyan, O., Topological indices for structure-activity correlations. In Steric Effects in Drug Design (Eds. M. Charton and I. Motoc), Topics Curr. Chem., 114 (1983) pp. 21–55, Springer, Berlin.
16. Sabljić, A. and Trinajstić, N., Acta Pharm.Jugosl., 31 (1981) 189.
17. Wiener, H., J. Am. Chem. Soc., 69 (1947) 17.
18. Balaban, A.T., Theor. Chim. Acta, 53 (1979) 355.
19. Bonchev, D., Balaban, A.T. and Randic, M., Internat. J. Quantum Chem., 1981, 19, 61. Erratum, ibid., 1982, 22, 441.
20. Hosoya, H., Bull. Chem. Soc. Jpn., 44 (1971) 2332.
21. Randić, M., J. Am. Chem. Soc., 97 (1975) 6609.
22. Randić, M., Chemometrics Intel. Lab. Syst. 10 (1991) 213. For newer literature in this field, see: Kezele, N., Klasinc, L., von Knop, J., Ivaniš, S. and Nikolić, S., Croat. Chem. Acta, 75 (2002) 651; Pompe, M., Chem. Phys. Lett. 404 (2005) 296.
23. Bonchev, D. and Trinajstić, N., J. Chem. Phys., 67 (1977) 4517.
24. Raychaudhury, C., Ray, S.K., Roy, R., Ghosh, J.J. and Basak, S.C., J. Comput. Chem., 5 (1984) 581.
25. Basak, S.C., Med. Sci. Res., 15 (1987) 605.
26. Basak, S.C., Magnuson, V.R., Niemi, G.J. and Regal, R.R., Discr. Appl. Math., 19 (1988) 17.
27. Basak, S.C., Harriss, D.C. and Magnuson, V.R. POLLY 2.3. University of Minnesota, Duluth, MN, 1998.
28. Balaban, A.T. and Balaban, T.S., J. Math. Chem., 8 (1991) 383.
29. Balaban, A.T., Chem. Phys. Lett., 89 (1982) 399.
30. Balaban, A.T., Pure Appl. Chem., 55 (1983) 199–206.
31. Balaban, A.T., Ionescu-Pallas, N. and Balaban, T.S., Math. Chem. (MATCH), 17 (1985) 121.
32. Balaban, A.T. and Quintas, L.V., J. Math. Chem. (MATCH), 14 (1983) 213.
33. Balaban, A.T., Math. Chem. (MATCH), 21 (1986) 115.
34. Ivanciuc, O., Ivanciuc, T. and Balaban, A.T., J. Chem. Inf. Comput. Sci., 38 (1998) 395.
35. Balaban, A.T., A comparison between various topological indices, particularly index J and Wiener's index W, In Rouvray, D.H. and King, R.B. (Eds.), Topology in Chemistry: Discrete Mathematics of Molecules. Horwood Publishing Ltd, Chichester, 2002, pp. 89–112.
36. Balaban, A.T., Mills, D. and Basak, S.C., MATCH Commun. Math. Comput. Chem., 45 (2002) 5.
37. Bertz, S.H., Discr. Appl. Math., 19 (1988) 41.
38. Vukičević, D. and Balaban, A.T., J. Internet Electron. J. Mol. Des. 4 (2005) 491.
39. Filip, P.A., Balaban, T.S. and Balaban, A.T., J. Math. Chem., 1 (1987) 61.
40. Programs POLLY and TRIPLET, University of Minnesota – Duluth, Natural Resources Research Institute.
41. Basak, S.C., Mills, D., Gute, B.D., Grunwald, G.D. and Balaban, A.T., J. Chem. Inf. Comput. Sci., 40 (2000) 891.
42. Basak, S.C., Gute, B.D. and Balaban, A.T., Croat. Chem. Acta, 77 (2004) 331.
43. Balaban, A.T. Mills, D. and Basak, S.C., Complexity of chemical graphs in terms of size, branching, and cyclicity (to be published).
44. Hall, L.H. and Kier, L.B., Quant. Struct.–Act. Relat., 9 (1990) 115.
45. Molconn-Z. 4.00, Hall Associates Consulting, 2 Davis Street, Quincy, MA and Molconn-Z 3.0S, Edusoft, LC, PO Box 1811, Ashland, VA, 23005.
46. Burden, F.R., J. Chem. Inf. Comput. Sci., 29 (1989) 225.
47. Pearlman, R.S. and Smith, K.M., J. Chem. Inf. Comput. Sci., 39 (1999) 28.
48. Stanton, D.T., J. Chem. Inf. Comput. Sci., 39 (1999) 11.
49. Baskin, I.I., Gordeeva, E.V., Devdaryanin, R.O., Zefirov, N.S., Palyulin, V.A. and Stankevich, M.I., Dokl. Akad. Nauk SSSR, 307(344) (1989) 613.
50. Gordeeva, E.V., Molchanova, M.S. and Zefirov, N.S., Tetrahedron Comput. Methodology, 3 (1990) 389.
51. Zefirov, N.S., Palyulin, V.A. and Radchenko, E.V., Dokl. Chem. 316 (1991) 43 (Dokl. Akad. Nauk SSSR, 316 (1991) 921).
52. Skvortsova, M.I., Stankevich, I.V. and Zefirov, N.S., Zh. Strukt. Khim., 33 (1992) 99.
53. Skvortsova, M.I., Baskin, I.I., Slovokhotova, O.L., Palyulin, V.A. and Zefirov, N.S., Dokl. Chem. 324 (1992) 103 (Dokl. Akad. Nauk SSSR, 324 (1992) 344).
54. Skvortsova, M.I., Baskin, I.I., Slovokhotova, O.L., Palyulin, V.A. and Zefirov, N.S., J. Chem. Inf. Comput. Sci., 33 (1993) 630.
55. Skvortsova, M.I., Baskin, I.I., Palyulin, V.A., Slovokhotova, O.L. and Zefirov, N.S., AIP Conf. Proc. (ECCC 1 Computational Chem.), 330 (1995) 486.

56. Baskin, I.I., Skvortsova, M.I., Stankevich, I.V. and Zefirov, N.S., J. Chem. Inf. Comput. Sci., 35 (1995) 527.
57. Skvortsova, M.I., Baskin, I.I., Slovokhotova, O.L., Palyulin, V.A. and Zefirov, N.S., Dokl. Chem. 346 (1996) 37 (Dokl. Akad. Nauk SSSR, 346 (1996) 497).
58. Skvortsova, M.I., Slovokhotova, O.L., Baskin, I.I., Palyulin, V.A. and Zefirov, N.S., Dokl. Akad. Nauk SSSR, 357 (1997) 72.
59. Skvortsova, M.I., Fedyaev, K.S., Palyulin, V.A. and Zefirov, N.S., Dokl. Chem., 379 (2001) 209.
60. Skvortsova, M.I., Fedyaev, K.S., Palyulin, V.A. and Zefirov, N.S., Internet Electronic J. Mol. Design, 2 (2003) 70.
61. Faulon, J.-L., J. Chem. Inf. Comput. Sci., 34 (1994) 1204.
62. Faulon, J.-L., J. Chem. Inf. Comput. Sci., 36 (1996) 731.
63. Faulon, J.-L. and Sault, A.G., J. Chem. Inf. Comput. Sci., 41 (2001) 894.
64. Faulon, J.-L., Churchwell, C.J. and Visco, D.P. Jr., J. Chem. Inf. Comput. Sci., 43 (2003) 721.
65. Churchwell, C.J., Rintoul, M.D., Martin, S., Visco, D.P. Jr., Kotu, A., Larson, R.S., Sillerud, L.O., Brown, D.C. and Faulon, J.-L., The signature molecular descriptor 3. Inverse quantitative structure-activity relationship of ICAM-1 inhibitory peptides. J. Mol. Graphics Modeling, 22 (2004) 263–273.
66. Faulon, J.-L., Collins, M.J. and Carr, R.D., J. Chem. Inf. Comput. Sci., 44 (2004) 427.
67. Visco, D.P. Jr., Pophale, R.S., Rintoul, M.D. and Faulon, J.-L., J. Mol. Graphics Modeling, 20 (2002) 429.
68. Faulon, J.-L., J. Chem. Inf. Comput. Sci., 38 (1998) 432.
69. Kier, L.B., Hall, L.H. and Frazer, J.W., J. Chem. Inf. Comput. Sci., 33 (1993) 143.
70. Kier, L.B., Hall, L.H. and Frazer, J.W., J. Chem. Inf. Comput. Sci., 33 (1993) 148.
71. Kier, L.B., Hall, L.H. and Frazer, J.W., J. Chem. Inf. Comput. Sci., 33 (1993) 598.
72. Kvasnička, V. and Pospíchal, J., J. Chem. Inf. Comput. Sci., 30 (1990) 99.
73. Kvasnička, V. and Pospíchal, J., Chemometrics Intell. Laboratory Systems, 11 (1991) 137–147.
74. Kvasnička, V. and Pospíchal, J., Chemometrics Intell. Laboratory Systems, 18 (1993) 171.
75. Kvasnička, V. and Pospíchal, J., J. Chem. Inf. Comput. Sci., 36 (1996) 516.
76. Klopman, G. and Raychaudhury, C., J. Chem. Inf. Comput. Sci., 30 (1990) 12.
77. Contreras, M.L. and Valdivia Rozas, R., J. Chem. Inf. Comput. Sci., 32 (1992) 323.
78. Goldman, D., Istrail S.A. and Piccolboni G.L., In: Proc. 11th ACM-SIAM Symposium on Discrete Algorithms, (2000) 275.
79. Raman, V.S. and Maranas, C.D., Comput. Chem. Eng., 22 (1998) 747.
80. Gupta, S., Singh, M. and Madan, A.K., J. Math. Analysis Applic., 275 (2002) 386.
81. Bajaj, S., Sambi, S.S. and Madan, A.K., Bioorg. Med. Chem., 12 (2004) 3695.
82. Sambi, S.S., Bajaj, S. and Madan, A.K., J. Mol. Struct. (Theochem), 684 (2004) 197.
83. Gupta, S., Singh, M. and Madan, A.K., J. Mol. Struct. (Theochem), 571 (2001) 147.
84. Bajaj, S., Sambi, S.S. and Madan, A.K., Bioorg. Med. Chem. Lett., 15 (2005) 467.
85. Bajaj, S., Sambi, S.S. and Madan, A.K., QSAR Comb. Sci., 23 (2004) 506.