ELSEVIER

Journal of Molecular Graphics and Modelling 20 (2001) 4–18

Journal of Molecular Graphics and Modelling

www.elsevier.com/locate/JMGM

Issues in representation of molecular structure The development of molecular connectivity

Lowell H. Hall [a,*], Lemont B. Kier [b]

[a] *Department of Chemistry, Eastern Nazarene College, Quincy, MA 02170, USA*

[b] *Department of Medicinal Chemistry, School of Pharmacy, Virginia Commonwealth University, Richmond, VA 23298, USA*

Received 24 October 2000; received in revised form 6 February 2001; accepted 7 February 2001

Abstract

Significant issues in the representation of molecular structure and the development of the molecular connectivity paradigm are presented. In the molecular connectivity paradigm, molecular structure is represented directly. Kier and Hall developed the method by creating ways to encode electronic information based on the paradigm developed from the Randić branching index. The simple and valence δ values were created to encode atomic and valence-state electronic information through counts of sigma, pi, and lone pair electrons. A family of indices was created to provide a wide range of structure information. The key aspects of the development are presented and discussed in such a way as to reveal, at least in part, the imaginative thinking involved in the process. Possible future roles for molecular connectivity chi indices are discussed.

Keywords: Molecular connectivity; Structure representation; Valence δ values; Chi indices; Structure–activity relationship; QSAR

1. Introduction

The year 1975 saw the publication of two papers that set the stage for a novel representation of molecular structure. Milan Randić described a skeletal branching index that correlated with three physical properties of alkanes [1]. Kier and coworkers showed that the branching index formalism could be applied to structures other than alkanes [2,3]. Based on these significant relationships, it was suggested that the encoded information was more broadly based and significant than skeletal branching. The branching index was renamed the molecular connectivity index [2]. In 1976, Kier and Hall published a paper [4] and a book [5] which developed the molecular connectivity idea into a full paradigm for the representation of molecular structure. Several basic issues involved in the development of molecular connectivity and representation of molecular structure remain important today and are the subject of this paper.

In the Randić paper, an alkane structure was converted to a hydrogen-suppressed graph. The count of neighbors in a graph has been called the vertex valence by mathematical graph theorists [5,6]. Randić combined vertex valencies into a pure number that encoded overall skeletal branching (Fig. 1). Each graph edge is characterized by the reciprocal square-root product of the vertex valencies. The branching index is the sum of the edge values. Table 1 includes the hexane isomer graphs, their branching index values, and the normal boiling point along with other property values. The remarkable aspect of the Randić paper is that a pure number, encoding alkane skeletal branching, also correlates with a physical property, boiling point. No use is made of calculations based on intermolecular forces or physical attributes, such as polarizability or volume.

The branching index is a quantitative whole molecule index that encoded specific attributes of structure. The algorithm resembled an additive scheme for chemical properties and appeared to have more potential for structure representation than the earlier graph-based formalisms [7]. The indices developed by Wiener [8], Platt [9], and Hosoya [10] were primarily counts of graph features. The Zagreb group was the first to propose indices which were directly based on the graph adjacency matrix [11]. (Some aspects of chemical graph theory are also discussed in papers by Randić [12], Bonchev [13], and Balaban [14] in this special issue.)

Two features of the branching index appeared open to fruitful investigation. First, although the branching index was suitable only for alkanes, ways could be imagined to modify the vertex valence to incorporate multiple bonding, heteroatoms and valence states. Although, explicit chemical information was missing from the algorithm, its emphasis on

* Corresponding author. Tel.: +1-617-745-3550.
E-mail addresses: hall@enc.edu (L.H. Hall), kier@hsc.vcu.edu (L.B. Kier).

1093-3263/01/$ – see front matter
PII: S1093-3263(01)00097-3

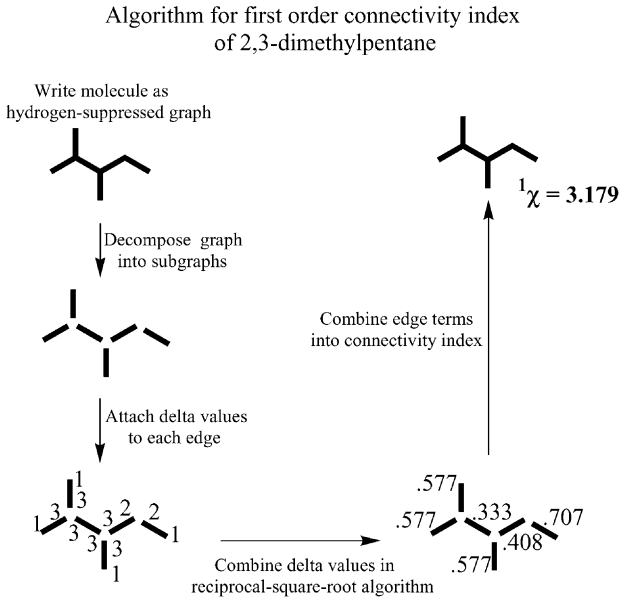


Fig. 1. Depiction of Randić branching scheme, showing the steps involved in the computation of the index for 2,3-dimethylpentane.

atoms and bonds provided a basis for development of chemical structure representation. Second, a single branching index was insufficient as a complete structure representation for a wide range of physical, chemical, and biological activities or for the wide range of structure features that occur in molecules. A method could be imagined for increasing the amount of structure information by extending the branching index formalism. The Randić paper was a first attempt at an interesting method; perhaps an investigation would reveal the possibilities [15].

2. Background of molecular connectivity

To investigate the significance of the branching index, Kier and Hall took the bold step of selecting a data set of nonspecific local anesthetics (Agin et al. [16] and Kier [17]) that includes a wide range of structures from methanol to quinine, a data set very different from the one used by Randić. When the anesthetic potency [log(minimum blocking concentration)] was plotted against the branching index [2], the smaller, simpler compounds fell along a tight straight line. However, when the larger, more potent polycyclic compounds were added to the plot, the set of points took on a more spread-out pattern. This scattered plot was less supportive of a relationship. It began to appear that the branching index was just a clever but simple idea that would not be adequate for the complexities of chemical and biological data.

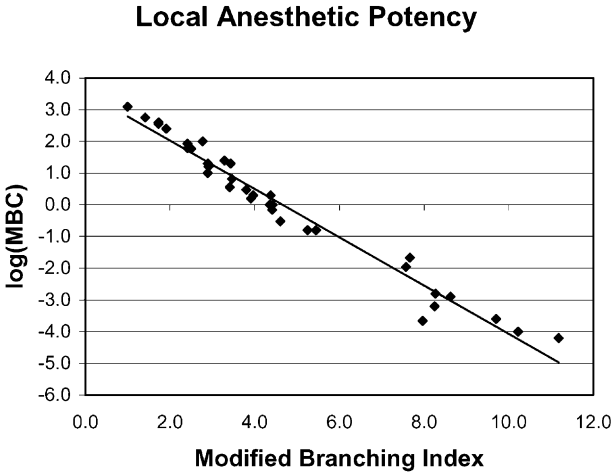


Fig. 2. Plot of observed anesthetic potency vs. modified branching index taken from [2].

Examination of the data suggested that an ad hoc algorithm modification might be helpful. An amount, 0.5, was deducted from the branching index for each ring. When this process was carried out systematically for all cyclic structures, an excellent plot and correlation were obtained. The relationship and accompanying statistics are as follows:

$$\log(\mathrm{MBC}) = -0.762\,(\text{modified branching index}) + 3.55,$$
$$r^2 = 0.97,\ s = 0.39,\ n = 36, \quad (1)$$

Table 1
Relation of properties to structure for hexane isomers[a]

	Branching index	Boiling point	Heat of formation	Liquid density	Molar refraction
	2.914	68.7 (1)	39.96 (5)	0.6594 (3)	29.907 (3)
	2.808	63.2 (2)	41.02 (4)	0.6643 (1)	29.802 (5)
	2.770	60.3 (3)	41.66 (2)	0.6532 (4)	29.946 (1)
	2.643	58.0 (4)	42.49 (3)	0.6616 (2)	29.810 (4)
	2.561	49.7 (5)	44.35 (1)	0.6492 (5)	29.935 (2)

[a] Number in parentheses is the rank order of the property values; for this data, the branching index inversely related to boiling point, directly related to heat of formation, but uncorrelated with density and molar refraction.

Given the diverse set of structures, these statistics indicate a significant relationship. The plot of calculated versus observed anesthetic potency is shown in Fig. 2.

Although, an ad hoc algorithm modification is not theoretically sound, it was believed justified in the very early stage of the investigation of a new algorithm. A potentially significant concept should not be eliminated at an early stage, because a reasonable but ad hoc modification was used. Subsequent investigations may reveal an approach that does not require ad hoc modification (see Sections 3 and 4) [5].

3. General considerations

It became clear that the success of the branching index was based on its representation of molecular structure. More than 'mere branching' was encoded. The molecular connectivity index was seen as a potentially novel approach to the broader and more significant concept of structure representation.

3.1. Nature of molecular structure

The development of the molecular connectivity paradigm progressed along two independent lines of thought. The first investigation considered ways in which molecular structure may be represented, raising a very basic question: what is molecular structure? For their working definition, Kier and Hall turned to the one given by Ernest L. Eliel in his book Stereochemistry of Carbon Compounds [18]: "The structure of a molecule is completely defined by the number and kind of atoms and the linkages between them.".

At its bedrock, molecular structure is the count of each atom, identified as its element, along with a description, enumeration or characterization of the set of connections between the atoms. To amplify this point, let us refer to a more general paradigm in science: the relation between form and function. In these terms, molecular structure is the form of the molecule. The measured physical properties are the function(s) of the molecule. Structure is not the set of physical properties for the compound, properties that arise from the structure. Structure represents the form of the substance which gives rise to its function, measured as its properties.

3.2. Relation of structure to property

The second important issue in the development of molecular connectivity as a paradigm deals with the question: how is structure related to property? In the usual modeling situation, a list of molecular structures is presented along with a corresponding list of property values (Table 2). The model should capture the parallel between structure variation and activity. The basic QSAR assumption is that activity may be related to molecular structure as a linear function and hundreds of examples have been published. It is also possible for QSAR equations to be nonlinear, as for example in the artificial neural network models. These concepts may also be represented diagrammatically in Fig. 3.

Table 2
The local anesthetic activity data set modeled by Kier et al. in their first connectivity paper [2]

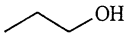
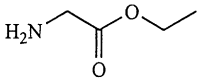
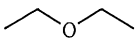
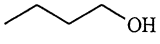
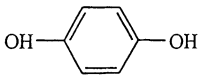
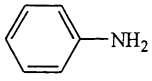
Number	Name	Molecular structure	log(MBC)[a]
1	Methanol	$H_3C–OH$	3.09
2	Ethanol		2.75
3	Acetone		2.60
4	Isopropanol		2.55
5	Propanol		2.40
6	Urethane		2.00
7	Diethyl ether		1.93
8	Butanol		1.78
9	Pyradine		1.77
10	Hydroquinone		1.40
11	Aniline		1.30

Table 2 (*Continued*)

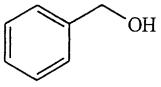
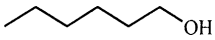
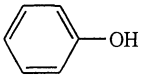
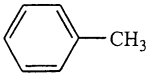
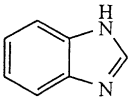
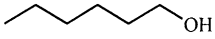
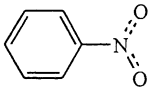
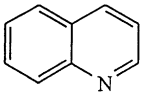
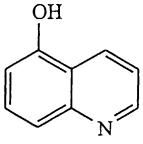
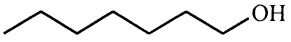
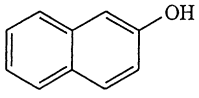
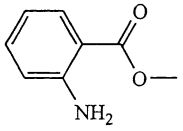
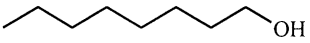
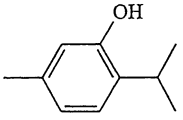
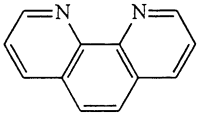
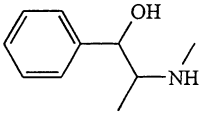
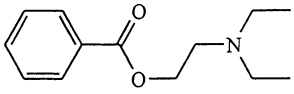
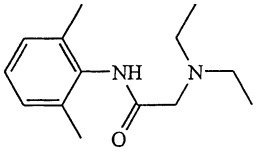
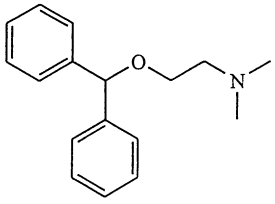
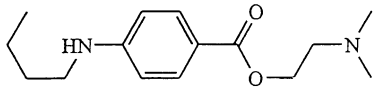
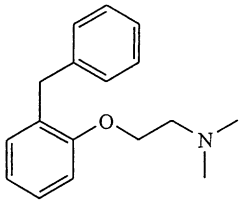
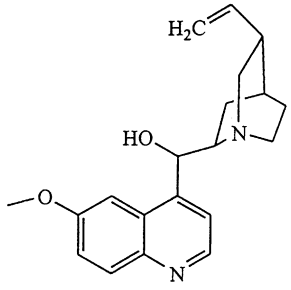
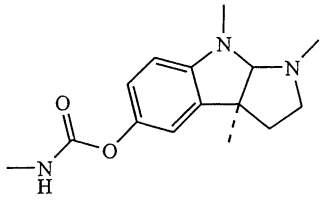
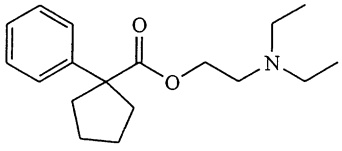
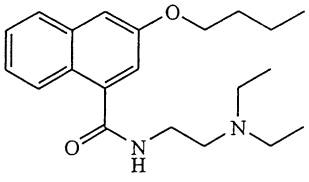
Number	Name	Molecular structure	log(MBC)[a]
12	Benzyl alcohol		1.30
13	Pentanol		1.20
14	Phenol		1.00
15	Toluene		1.00
16	Benzimdazole		0.81
17	Hexanol		0.56
18	Nitrobenzene		0.47
19	Quinoline		0.30
20	8-Hydroxyquinoline		0.30
21	Heptanol		0.20
22	2-Naphthol		0.00
23	Methyl anthranilate		0.00
24	Octanol		−0.16
25	Thymol		−0.52
26	*o*-Phenanthroline		−0.80
27	Ephedrine		−0.80
28	Procaine		−1.67
29	Lidocaine		−1.96

Table 2 (*Continued*)

Number	Name	Molecular structure	log(MBC)[a]
30	Diphenhydramine		−2.80
31	Tetracaine		−2.90
32	Phenyltoloxamine		−3.20
33	Quinine		−3 60
34	Physostigmine		−3.66
35	Caramiphen		−4.00
36	Dibucaine		−4.20

[a] Minimum blocking concentration.

QSAR deals primarily with noncovalent properties, i.e. properties arising from interactions in which covalent bonds are neither formed nor broken. These interactions are manifestations of intermolecular forces. The critical aspect of the structure–property relationship is the adequate representation of the molecular structure so that statistical analysis may be used to establish the QSAR/QSPR models. These concepts have been discussed in several books [5,19–25].

4. Development of molecular connectivity

The Randić algorithm was based on the vertex valence. The use of the hydrogen-suppressed graph raised a basic question: are the hydrogen atoms actually suppressed? Only skeletal atoms are counted. This feature was disturbing, since hydrogen atoms are important in intermolecular interactions; this feature was also attractive because the graph

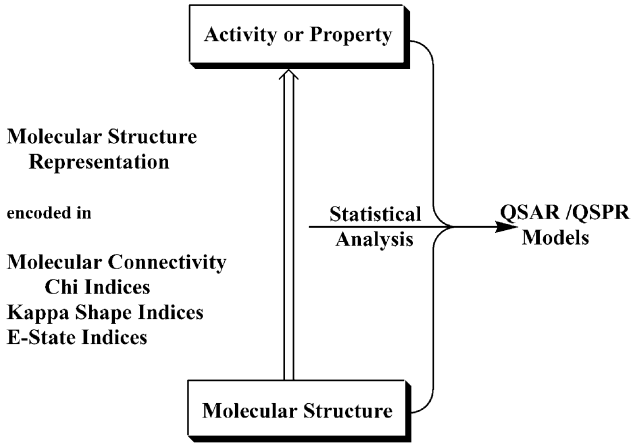


Fig. 3. Depiction of relation between molecular structure (form) and measured properties (function).

is a succinct and economical expression of structure. Recently, it has been suggested that a more accurate name for the graph is hydrogen-implied graph [15].

Structure information may be expressed in a hierarchy with the molecular formula and its atom counts at the lowest level, groups and bond types at a higher level, and the hydrogen-suppressed graph-based information at an even higher level. Basak et al. introduced the useful terms topostructural, topological, and topochemical [26–28].

4.1. Creation of the valence δ value

Vertex valence, the number of skeletal neighbors for each vertex, is not an explicit expression of electronic structure. Kier and Hall tried to imagine a connection between vertex valence and electronic structure that could relate to electron counts.

Number of carbon neighbors (vertex valence, v) for a carbon atom in an alkane is given simply as $v = 4 - h$, in which h is the count of bonded hydrogens. The number 4 stands for number of valence electrons in carbon as well as the number of electrons assigned to sigma orbitals. That equality was taken as the basis for a new definition:

$$v \rightarrow \delta = \sigma - h.$$

In this manner, the term 'δ value' was introduced and δ given as a symbol for use in the molecular connectivity formalism [3]. This reinterpretation of vertex valence was an important step from mathematics to chemistry, i.e. moving from simple counting of abstract objects to use of explicit chemical electronic information.

Early in the investigations, an index was needed which encodes chemical element and valence states. Empirical valence δ were developed, derived by regression analysis from physical properties such as molar refraction [5], but a theoretical basis was needed. Kier and Hall tried to imagine what could parallel the sigma electron count and at the same time deal with heteroatoms in a meaningful way. It appeared that the count of all valence electrons might be useful. Further, there must be a role for the total number of electrons in an atom to relate to factors such as atomic size, polarizability, and electronegativity.

Since the simple δ value is restricted to the sigma electron count, a valence δ could be developed which includes the total valence electron count [4]. Those relations are stated as follows:

$$\delta = \sigma - h, \tag{2}$$

$$\delta^v = Z^v - h = \sigma + \pi + n - h. \tag{3}$$

A selection of simple and valence δ values is shown in Table 3 for second row atoms. Simple and valence δ values are identical for carbon sp^3 atoms, because $\sigma = Z^v$. For other atoms and valence states, $Z^v > \sigma$ and $\delta^v > \delta$, expressing the presence of pi and lone pair electrons in the valence state. These equations also reveal that hydrogen atom count is actually included in the formalism developed from the graph. This realization gives further support to using the term hydrogen-implied rather than hydrogen-suppressed graph.

The expression in Eq. (3) is satisfactory for second row atoms, but the electron core must be taken explicitly in account for higher row atoms. The general aspects of the valence δ were presented in a 1983 paper [29]. A denominator was introduced into the valence δ expression to account for the effect of core electrons as follows:

$$\delta^v = \frac{Z^v - h}{Z - Z^v - 1} = \frac{\sigma + \pi + n - h}{Z - Z^v - 1}. \tag{4}$$

Selected values are given in Table 3. In this relation, Z is the atomic number. The denominator, $Z - Z^v - 1$, reduces to 1 for second row atoms. Larger atoms have smaller δ^v values and make a larger contribution to chi indices. This general expression also gave values that were similar to the empirical valence δ values introduced earlier.

4.2. Significance of valence δ value

The significance of the creation of a general expression for the δ values is the development of a whole paradigm,

Table 3
Selected examples to illustrate the molecular connectivity simple and valence δ values

	$-CH_3$	$-CH_2-$	$=CH_2$	$-NH_2$	$-NH-$	$>N-$	$-OH$	$-O-$	$=O$	$-F$	$-Cl$	$-Br$	$-I$	$-S-$
δ	1	2	1	1	2	3	1	2	1	1	1	1	1	2
δ^v	1	2	2	3	4	5	5	6	6	7	0.78	0.26	0.16	0.56

advancing molecular connectivity from a measure of graph branching for alkanes to a formalism for representation of organic molecule structure.

In the further development of the molecular connectivity paradigm, it was important to explore any connection between the connectivity δ values and fundamental aspects of structure. Pi and lone pair electrons are more exposed and more interactive than the sigma network electrons, and hence more important to noncovalent intermolecular interaction. Because the sigma, pi, and lone pair (n) electron counts are included in both the simple and valence δ values, it was of interest to investigate their relationships. It is clear from Eqs. (2) and (3) that the difference in δ values held significant electron count information:

$$\delta^v - \delta = \pi + n. \tag{5}$$

This simple relation makes readily accessible the count of these electrons that are usually more significant in molecular interactions. The electrons assigned to pi molecular orbitals can arise from atomic p, d, or f orbitals.

4.3. Valence-state electronegativity

Electronegativity is such a useful concept that a new valence-state electronegativity was defined as essentially the count of pi and lone pair electrons. These electrons are ineffective in shielding the nuclear charge for atoms bonded in a covalent molecule. In a 1981 paper, this idea was advanced. A strong relation was demonstrated between the topological valence-state electronegativity and experimental values, the Mulliken–Jaffe valence-state electronegativity (based on experimental valence-state ionization potentials (IP) and electron affinities [30]). The general definition of Kier–Hall valence-state electronegativity, X_{KH}, is as follows:

$$X_{\mathrm{KH}} \rightarrow \frac{\delta^v - \delta}{N^2} = \frac{\pi + n}{N^2}, \tag{6}$$

where N is the principal quantum number of the valence electrons for the element. The correlation relation obtained is as follows:

$$X_{\mathrm{MJ}} = \frac{7.99(\delta^v - \delta)}{N^2} - 7.07,$$
$$r^2 = 0.98,\ s = 0.48,\ F = 600,\ n = 19. \tag{7}$$

This valence-state electronegativity is in general agreement with other electronegativity scales, illustrated as follows: $X_{\mathrm{KH}}(\text{–OH}) > X_{\mathrm{KH}}(\text{–NH}_2) > X_{\mathrm{KH}}(\text{–CH}_3)$. Furthermore, other effects in organic molecules are also described very well: $X_{\mathrm{KH}}(\text{C}{\equiv}\text{H}) > X_{\mathrm{KH}}(=\text{CH}_2) > X_{\mathrm{KH}}(\text{–CH}_3)$.

This electronegativity definition is useful in itself, but it also led to a big step further in the development of topological representations of molecular structure. The Kier–Hall valence-state electronegativity became the basis of the intrinsic state that is an integral part of the electrotopological state formalism, the subject of a recent book [19]. Whereas, the chi indices are whole molecule descriptors, E-state indices characterize each individual atom and are also expressed as atom types.

4.4. Formal definition of first-order molecular connectivity

Based on the general definition of the simple and valence δ values, the branching index could be transformed into the molecular connectivity chi index (of the first order):

$$^1\chi = \sum_k (\delta_i \delta_j)_k^{-1/2} = \sum_k c_k,$$
only sigma electrons, treating all atoms as carbon sp^3, (8)

$$^1\chi^v = \sum_k (\delta_i^v \delta_j^v)_k^{-1/2},$$
valence electrons for all elements. (9)

Some investigators have referred to the valence indices as ‘valence-corrected’. Based on the development just given, it can be clearly seen that the term ‘valence-corrected’ is not appropriate. The valence indices are simply what they are called, chi indices defined for all electrons in all elements, not corrected simple indices.

4.5. The δ chi index

The simple chi index is limited to sigma electrons and the valence index includes all the valence electrons. Hence, their difference called the delta chi index tends to emphasize the pi and lone pair electrons. The delta chi index of first order was defined as follows:

$$\Delta^1\chi = {}^1\chi - {}^1\chi^v. \tag{10}$$

For alkanes and saturated bonds of molecules, $^1\chi = {}^1\chi^v$, then $\Delta^1\chi^v = 0$ for alkanes and for alkyl parts of a molecule. The role of the delta chi index was examined for QSAR type models. One important place where the role of pi and lone pair electrons is very significant is in the ionization potential, since pi and lone pair electrons tend to contribute significantly to the HOMO state. The ionization potentials (in eV) were considered for a set of 24 alkyl amines, alcohols, and ethers [31]:

$$\mathrm{IP} = 5.364\,\Delta^0\chi + 6.341\,\Delta^1\chi + 1.517\,\Delta^2\chi + 4.243,$$
$$r = 0.993,\ s = 0.12,\ F = 461, \tag{11}$$

where the superscript refers to the number of atoms in the corresponding subgraph. See below for the discussion of extended chi indices. The $\Delta^0\chi$ term, which considers one atom at a time, is exactly zero for all atoms but the heteroatoms, nitrogen and oxygen. The ionization potential is the energy of the HOMO, highly localized on the heteroatom for these

molecules. The $\Delta^0\chi$ term represents this structure information approximately. The $\Delta^1\chi$ term represents graph edges; only those between a heteroatom and the alpha-carbon are nonzero. This term represents differentiation among primary, secondary and tertiary amines as well as between alcohols and ethers. The alpha-carbon atoms influence the HOMO by altering the electron density on the heteroatom; the $\Delta^1\chi$ term represents this structure information. Finally, the $\Delta^2\chi$ term adds structure information representing the refinement of structure in carbon atoms alpha and beta to the heteroatoms, encoding branching in the alkyl portions of the molecules. The statistics for Eq. (11) are excellent, providing both a useful predictive model as well as some insight into the structure influences on ionization potential for these compounds.

The delta chi indices were also used to model bioconcentration factor (BCF) of halogenated organics. Both the sum index ($\sum {}^0\chi = {}^0\chi + {}^0\chi^v$) and delta chi index ($\Delta^0\chi = {}^0\chi - {}^0\chi^v$) of zero-order chi indices appear in the model equation as shown [32]:

$$\begin{aligned}\log(\mathrm{BCF}) = {} & 0.795 \sum {}^0\chi - 0.0170 \left(\sum {}^0\chi\right)^2 \\ & -0.530\,\Delta^0\chi - 0.787(\Delta^0\chi)^2 \\ & +0.0632 \left(\sum {}^0\chi\,\Delta^0\chi\right) - 4.735, \\ & r = 0.984,\ s = 0.22,\ F = 88,\ n = 20. \end{aligned} \quad (12)$$

The data points lie across the surface of a generalized parabola that requires the linear, quadratic and cross-terms in the algebraic equation. This data set includes chlorinated hydrocarbons, benzenes, biphenyl and diphenyloxides. In addition DDT, heptachlor and dieldrin also fit the model well. The optimum in the surface is near the values for tetrachlorobiphenyl. Other applications of the delta chi indices are listed below.

4.6. Meaning of molecular connectivity

By 1986, the utility of ${}^1\chi$ was clearly established in many papers [5,20,31–42]. It became important to establish the fundamental meaning of chi indices. The index, ${}^1\chi$, is a direct representation of molecular structure that encodes degree of skeletal branching. The reinterpretation of vertex valence as sigma electron count indicated that the connectivity index also encodes basic aspects of electronic structure information. The first-order chi index is not a surrogate for some property any more than a correlation of boiling point with number of atoms means that number of atoms is a surrogate for boiling point. N is simply a count of atoms or bonds, ${}^1\chi$ a weighted count of skeletal bonds. In this sense, ${}^1\chi$ is a direct representation of molecular structure.

The ${}^1\chi$ index is calculated through an algorithm that is based directly and entirely on the skeletal structure. No intermediary property is employed or necessary. The connectivity index encodes certain aspects of skeletal structure information, to the extent that the encoded information also parallels a property, a model based on the connectivity index describes the structure–property relationship. As Adamson put it, models based on topological indices can be called "mechanism-free" [43,44]. Thus, models based on the connectivity index directly are not mechanism-biased.

In a recent development, Kier and Hall have shown that the connectivity index possesses an even more fundamental meaning in terms of intermolecular accessibility [45,46]. This development is discussed in the second paper in this special issue.

4.7. Extending structure information in the extended chi indices

Development of a means for encoding explicit chemical electronic information in the δ values allowed Kier and Hall to turn their attention to a second issue, the need for more than a single index to encode structure information. Based on the general experience in property variation with structure variation, more than a single index is necessary to encode the rich and widely varying structure information in a set of molecules.

Consider for example that the rank ordering of hexane isomer boiling points is not the same as for molar refraction, as shown in Table 1. For this brief selection of data, boiling point is directly related to the connectivity index, whereas enthalpy of formation is inversely related. Neither liquid density nor molar refraction is well correlated with the branching index. Many other examples clearly indicate that more than one structure factor is involved in structure–property relationships in general. Sound statistical models usually require more than a single variable in multiple linear regression [4].

This information provided an opportunity to think imaginatively. The ${}^1\chi$ index arises from decomposition of the graph into subgraphs of bonds, paths of length one edge. Decomposition into larger subgraphs could also be considered. Use of subgraphs composed of several edges was not a new concept for mathematical graph theorists [5,6]; however, using such fragments of molecular structure in a structure-encoding algorithm was novel.

Fig. 4 depicts the connected subgraphs of the isopentane skeleton. Subgraphs may consist of a single atom, a single edge or a set of connected edges (path) in which no vertex is included twice. Other types of subgraphs may be used as depicted in the figure, including rings, not shown [47].

4.8. General definition of chi indices

Greater sensitivity to structure variation (than obtained by mere counting) can be achieved by adopting an algorithm similar to the branching index algorithm. Higher-order or extended chi indices can be defined in an analogous manner as follows:

$$ {}^m\chi_t = \sum_1 \prod_i (\delta_i)_1^{-1/2}, $$

sum over all type t subgraphs with m edges. (13)

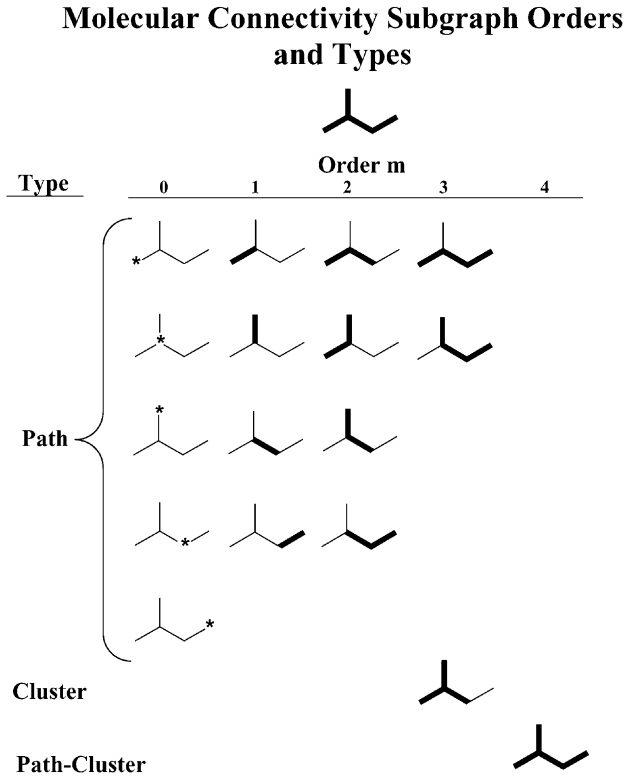


Fig. 4. The subgraphs of the isopentane skeleton as an illustration of subgraphs used in molecular connectivity.

A parallel set of valence chi indices may also be developed as follows:

$$^{m}\chi_{t}^{v} = \sum_{1}\prod_{i}(\delta_{i}^{v})_{1}^{-1/2},$$

sum over all type t subgraphs with m edges using the valence δ values. (14)

This formalism provides a set of chi indices to encode a wide range of structure features. A set of histograms can be used to represent heptane isomers and indicate the characterization of their various skeletons with the set of chi indices [5].

4.9. Applications of chi indices

So far in our discussion, we have dealt primarily with an index based on graph edges, that is the first-order index. We can give brief illustration of the family of chi indices, including extended chi indices.

4.9.1. Nonspecific local anesthetics

As a first example, let us refer back to the nonspecific local anesthetics. The model, published in 1975, made use of an ad hoc algorithm modification to accommodate ring structures. The family of chi indices provides models in which that modification is not necessary. Correlation with the zero-order valence chi index gives slightly better statistics than given in Eq. (1):

$$\log(\mathrm{MBC}) = -0.5460\,^{0}\chi^{v} + 3.66,$$
$$r^2 = 0.97,\ s = 0.37,\ n = 36. \quad (15)$$

The wide range of applications of the chi indices has indicated that an ad hoc algorithm modification is not necessary.

4.9.2. Simple chi structure space

To determine whether the family of chi indices actually extends the range of structure information, let us consider the third-order path chi index along with the first-order index. The first-order index ($^{1}\chi$) encodes degree of branching; the index decreases as the degree of skeletal branching increases. This index gives little information about the organization of branching within the skeleton. Consider Fig. 5 to explore the information encoded by both $^{1}\chi$ and the third-order path index $^{3}\chi_{p}$; it includes acyclic, monocyclic, and bicyclic octane isomers [48].

The arrangement of graphs in Fig. 5 clearly indicates that structure information is encoded in $^{1}\chi$. As the index value decreases, the degree of skeletal branching increases, whether dealing with acyclic or cyclic skeletons. If only edge counts were used in place of $^{1}\chi$, all the graphs would fall onto three points: one corresponding to seven edges for acyclic, one with eight points for monocyclic, etc. The reciprocal-square-root weighting scheme provides significantly more structure information. The $^{1}\chi$ index provides little information on the arrangement of branch points. However, the octane skeletons are spread out horizontally in a chemically meaningful manner. An important graph feature encoded in $^{3}\chi_{p}$ is adjacency of branch points. The greater adjacency in branching, the further to the right the graphs are located. Adjacency of branch points influences both the ^{3}p path count as well as the actual value of δ values in the path term. Further, diagonal lines represent constant number of branching points.

Based on this brief analysis, it is clear that these two chi indices provide a chemically meaningful organization of octane skeletons, suggesting that simple chi indices provide the basis for a structure space in which molecular skeletons are arrayed in a useful manner. Much more useful information is encoded than in the simple path counts. Simple chi indices provide the basis for database searching when molecular skeletons are the important issues [48].

4.9.3. Liquid density

Since chemically meaningful information is encoded in the extended chi indices, it is expected that useful correlation models would be obtained by including the extended chi indices. In the first book on connectivity, many studies involving the extended indices are given. As one example, consider a model for the density of liquid alkanes [5]. For 82 alkanes, the following model was developed for liquid density:

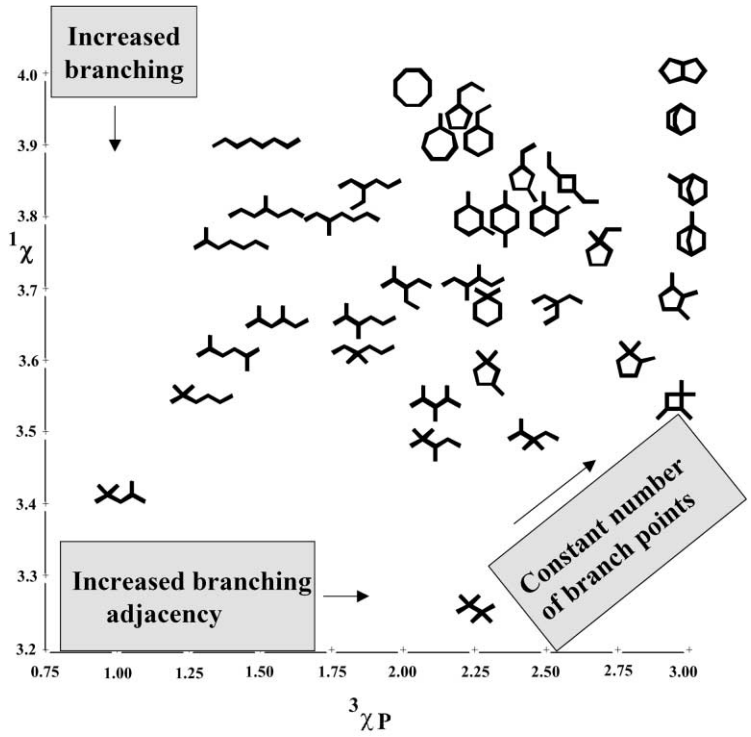


Fig. 5. Plot of acyclic, monocyclic, and bicyclic octanes with the first-order simple connectivity index vs. the third-order simple chi index.

$$d = 0.7348 - \frac{0.2929}{{}^1\chi} + 0.0030\,{}^3\chi_{\mathrm{p}},$$
$$r^2 = 0.98,\ s = 0.0046,\ n = 82. \qquad (16)$$

The third-order chi path index adds significantly to the statistical significance of the model; the correlation coefficient with the single variable $1/{}^1\chi$ is $r = 0.903$. Also note that this application is among the first to make use of topological indices in a nonlinear fashion. Density models for other classes of compounds were also developed [5].

4.9.4. Other applications

Based on these findings with the extended chi indices, it is clear that the set of principal components of simple chi indices may organize a chemical database according to similarity of molecular skeletons. Basak et al. was the first to publish significant use of structure space creation with the use of principal components [49,50]. Many authors have made significant contributions to the application of the molecular connectivity method. Seybold and coworkers showed models based on the chi indices for boiling point and analyzed the properties of halocarbons [51–55]. Balaban and Balagan examined the use of real number vertex invariants and the degeneracy of topological indices [56–58]. Bonchev and coworkers developed the related area of information theoretic indices [59,60]. Recently, Bonchev has developed the concept of overall connectivity [61]. Galvez et al. [62,63,65] and Garcia-Domenech et al. [64] have made significant use of the delta chi indices in drug design and illustrated the method in several papers. Pogliani has shown in a series of papers how to construct useful linear combination of molecular connectivity indices, LCCI–MCI [66–70]. Kier and Hall assigned a value to each atom in a alkane by partitioning ${}^0\chi$, ${}^1\chi$, and ${}^2\chi$ values. They found that these atom-based values relate well to the electron density of alkyl groups. This approach is also an early precursor to atom level indices in the E-state formalism [71].

4.9.5. QSAR with higher-order chi indices

A set of QSAR models was developed for for phenylpropylether activity against *Trichophyton mentagrophytes* and for the antiviral activity of a set of benzimdazoles [72]. These two examples revealed an approach to structure representation with the chi indices which involves decomposition of the indices into their subgraph terms. Subgraph terms for specific parts of the molecules were found to give regression models of high statistical quality. Thus, specific parts of the molecule were highlighted as important.

For the phenylpropylethers, the following model was found:

$$\log\left(\frac{1}{c}\right) = 1.30^{1}\chi - 2.703^{3}\chi_{\mathrm{p}} + 2.74^{4}\chi^{v}_{\mathrm{PC}} + 0.01,$$
$$r^2 = 0.91,\ s = 0.15,\ F = 84,\ n = 28. \quad (17)$$

A Hansch type analysis found a less satisfactory correlation (only after the deletion of two compounds which were large outliers): $r = 0.83$, $s = 0.22$ [5]. In the data set, two positions on the propyl group are substituted with one or two –OH groups. The phenyl ring is also substituted by various numbers and combinations of methyl and chloro groups. The third- and fourth-order terms were partitioned into localized parts of the molecules by taking the subgraph terms for those parts. A new term was created which was the sum of the path-three chi index subgraph terms which contained any of the –OH groups called $^{3}S_{\mathrm{p}}$. A term was also created for fourth-order path–cluster subgraphs that include the para-substituent on the phenyl ring, called $^{4}\mathcal{S}^{v}_{\mathrm{PC}}$. When the $^{3}S_{\mathrm{p}}$ and $^{4}\mathcal{S}^{v}_{\mathrm{PC}}$ terms replaced the corresponding chi indices in the regression model, the statistics did not degrade much, $r^2 = 0.85$ compared to $r^2 = 0.91$ for the full chi indices. These results suggest that the dihydroxy region and the para-region have a strong influence on the activity.

In the second study, it was found that only higher-order chi indices would yield a satisfactory regression model for the antiviral activity of benzimidazoles [72]:

$$\log\left(\frac{1}{c}\right) = 1.89^{6}\chi_{\mathrm{p}} - 0.677^{4}\chi^{v}_{\mathrm{p}} + 1.04,$$
$$r^2 = 0.93,\ s = 0.14,\ F = 87,\ n = 15. \quad (18)$$

Regression with the $^{4}\chi^{v}_{\mathrm{p}}$ index alone gives $r^2 = 0.93$, $s = 0.17$. A Hansch type analysis with the π parameter gives $r^2 = 0.90$; there is no discrimination among the substitution positions. Discrimination was demonstrated among ring positions based on a subgraph analysis similar to that for the phenylpropylethers.

4.10. The dimensionality question

In recent years, the range of approaches to representation of molecular structure has presented the chemical community with an array of structure descriptors ranging from quantum mechanical, to topological indices, to group additive feature counts, to calculated estimates of physical property values. In a natural attempt to classify these descriptors and methods, one approach refers to descriptors in 'dimensionality' terms. Simple structure feature counts such as number of atoms have been called one-dimensional (1D). Topological descriptors have been called two-dimensional (2D); those derived from quantum mechanics or crystal structure data have been called three-dimensional (3D). It appears that some of these adjectives are inappropriate. Fundamentally, the adjective 'dimensional' comes from a coordinate system, such as Cartesian or generalized space. In the case of the molecular connectivity chi indices, their origin is not a coordinate system but rather the adjacency matrix. The chi indices arise from the chemical graph. From this perspective, it is inappropriate to refer to chi indices in dimensional terms, whereas 3D does appear appropriate for quantum-mechanically derived quantities or parameters obtained from crystal structure information.

Chi indices and related quantities are derived from the mathematical graph that has been invested with explicit electronic information. The graph is a set of objects (vertices, atoms) and a set of relationships among the objects. For example, the atoms can be related to the count of sigma, pi and lone pair electrons as in the simple and valence δ values. Atoms can also be related to the set of skeletal paths as in the extended chi indices. For this reason, a dimensionality adjective does not accurately describe the nature of the indices. The terms recommended are topological or topochemical [26–28,42].

4.11. Interpretability of chi indices

A model based on the chi indices may be interpreted directly in terms of significant aspects of molecular structure. Using basic structure concepts, such as degree of branching, molecular size, heteroatom states, cyclization and related terms, it is possible to gain structure insight from the variables in the model. These terms are not synonymous with some of the terms used in physical organic chemistry for one type of molecule description. Terms such as lipophilicity, electronic character, polarity and others are often used. The structure-based terms and the property-based terms are not in conflict. However, the use of structure-based terms relates more directly to the information required to describe the structure of molecules whose synthesis is required.

A QSAR model developed on the toxicity of phenols to fish [73] can serve as an illustration:

$$\mathrm{pLC}_{50} = 0.205^{1}\chi + 0.906^{3}\chi^{v}_{\mathrm{p}} + 1.786,$$
$$r^2 = 0.87,\ s = 0.30,\ F = 75,\ n = 25. \quad (19)$$

This paper also examined the statistical and predictive significance of the model. The independent variables were replaced with random numbers and the model-building process repeated. No statistically significant models were found. The toxicity values were also randomized; no significant models were found. Observations were randomly deleted (three per run) until every observation had been eliminated once. The coefficients and the regression statistics were found to be very stable. Based on the significance of the model, structure interpretations of the indices were developed.

The first-order chi index, $^{1}\chi$, contributes 40% of the calculated toxicity. It encodes molecule size and skeletal degree of branching, independent of the nature of the atoms. With a positive coefficient, the model indicates that toxicity generally increase with size. It is clear, however, that toxicity would not be expected to increase infinitely with size; the stated size dependence can only be applied to the range of the present data set. If data were available for much larger

molecules, it is expected that toxicity would decrease eventually with size. Quadratic dependence on $^1\chi$ and $^1\chi^v$ has been reported for chi indices [5,20,32,48].

The second variable, $^3\chi_p^v$, accounts for 60% of the calculated toxicity. The path-three index is larger for adjacent branch points than for separated branch points. For example, the 3,4-dimethyl derivative has a larger toxicity than does the 2,4-dimethyl compound. In general, toxicity is greater for adjacent branching. Further, the path-three valence index $^3\chi_p^v$ has larger values for bromine than for chlorine, illustrated by the toxicity of the 2,4,6-tribromo-compound (larger than for the 2,4,6-trichlorophenol). The model indicates that toxicity is decreased for smaller, less-branched compounds, with nonadjacent substitution and lower atomic weight halogens. As a result, the investigator has the information necessary for further molecule design.

4.12. Computation

It became clear very early in the development of molecular connectivity that it was necessary to create computer software to obtain all the possible subgraphs and to input various structure formats. In the late 1970s, Hall developed the CFUNC software. This primitive FORTRAN program has been supplanted by significantly upgraded software packages and now exists in the form of Molconn-Z [74]. Molecular connectivity chi indices are now available in software from most chemical software houses.

4.13. Future of chi indices

The future is difficult to predict for an area of active involvement by investigators in several fields. Nonetheless, we can indicate three of the aspects of the use of chi indices.

1. First, chi indices are being used as part of the set of variables considered for development of QSAR models rather than the only structure descriptors available. For example, Hall and Maw are currently developing a model for binding of HIV protease inhibitors [75]. The data are taken from Holloway et al. [76]. The current model is as follows:

$$\begin{aligned} \mathrm{pIC}_{50} &= 0.743\,{}^1\chi^v + 0.703\,{}^2\kappa_\alpha + 0.176\mathrm{SH}^\mathrm{T}(\mathrm{other}) \\ &\quad + 0.693\mathrm{Hs}(\mathrm{C}_{11}) + 0.517\mathrm{SH}^\mathrm{T}(\mathrm{HB_d}) - 2.915, \\ r^2 &= 0.85,\ s = 0.62,\ n = 32,\ r^2_{\mathrm{press}} = 0.73. \end{aligned} \quad (20)$$

   The $^1\chi^v$ index accounts for an average of 33% of the calculated binding for the data set. In this sense, $^1\chi^v$ plays a significant role in the model, but does not encode all the structure information necessary for a good model. The other indices are as follows: $^2\kappa_\alpha$ is the second-order kappa shape index [32], SH^T(other) the atom-type hydrogen E-state index for nonpolar atoms, $\mathrm{SH}^\mathrm{T}(\mathrm{HB_d})$ the atom-type hydrogen E-state index for hydrogen bond donors, $\mathrm{Hs}(\mathrm{C}_{11})$ the atom-level E-state index for a particular atom in the common skeleton of the data set [19]. Inclusion of one or more chi indices along with other topological indices is common as in a model for the toxicity of a set of 50 amide herbicides in which $^2\chi^v$ is a significant variable [77].
2. A second area for continued and future applications of chi indices is in the creation of models for large data sets. There are a few properties for which much experimental data has been accumulated, including log *P*, boiling point, melting point, pK_a. Recently, Parham et al. have developed a model for log *P* based on 12,943 compounds [78]. The model is based on 108 structure descriptors, including molecular connectivity chi indices, kappa shape indices, and atom-type E-state.

   The modeling uses artificial neural networks. Variables were selected within the nonlinear neural network context rather than with linear methods. Ten percent of the data was set aside for external validation. The remainder was randomly divided into subsets of 10% many times so that each compound appeared in a test set two times and in a training set eight times. As a result, each compound log *P* value was predicted several times. A consensus prediction was developed. For all the training sets, $r^2 = 0.959$, MAE = 0.30, $n = 10{,}353$. For the test sets, $r^2 = 0.927$, MAE = 0.36, $n = 10{,}353$. For the external validation set, $r^2 = 0.935$, MAE = 0.35, $n = 1258$. This model may be evaluated at the website: www.lop.com. Parham et al. report that results for compounds that are not very similar to the original data set indicate that the model has good predictive power [78].

   The combination of topological indices, artificial neutral networks and large data sets has been extended to the modeling of aqueous solubility. This model may also be evaluated on www.logP.com.
3. A third area of future use of the molecular connectivity chi indices is in the inverse QSAR problem. When a good QSAR equation is developed, based on topological indices, can that mathematical equation be inverted, in the mathematical sense, to indicate those structures that would possess a specified range of property/activity values? The basic issues were addressed in a series of four papers [79–82].

   The approach focused on graph primitives, path counts, and vertex degree counts. Given a particular chi index from a regression model, say $^2\chi$, the corresponding path count, ^2p, could be obtained for a specified value of $^2\chi$. If the set of vertex degrees, ^mD, could be obtained, then the corresponding graph(s) could be constructed. Therefore, the basic problem was to find a relationship between path counts and vertex degree counts. The first two papers in this series gave derivations of a set of information transfer equations. The first paper also demonstrated an algorithm for the complete inversion process, using molar volume of alkanes to illustrate the method. A third paper extended the equations to include the path-three counts,

^3p, which also involved the counts of edge types [81]. This development extended the range of applicability of the method. The fourth paper demonstrated the inverse QSAR method on a biological data set [82].

In a continuation of the inverse QSAR methodology, Fisk and Hall [83] developed an algorithm for creation of vertex degree sets, ^mD, derived from only atom count and ring count, N and R, involving detection of multigraphs and pseudographs. The algorithm was illustrated with many examples.

Future work in the inverse QSAR problem will probably require incorporation of the E-state formalism that permits focus on individual atoms. The combination of the whole molecule character of the chi indices with the individual atom nature of the E-state is the basis for a powerful approach.

5. Conclusions

A revolution has been occurring in property and activity modeling since it was realized that structure is more than geometry or conformation and different from physicochemical properties. Molecular connectivity is a paradigm in which molecular structure is represented directly. This paradigm was developed from the Randić algorithm for a branching index so that electronic information was encoded. A family of indices was developed to provide for a wide range of structure information. Atomic and valence-state electronic information is encoded in the simple and valence δ values that include counts of sigma, pi, and lone pair electrons.

Molecular structure may be viewed as a collection of information that can be synthesized into a mathematical model of a property. In this sense, direct structure–activity models have been developed. We are only beginning to tap into that great reservoir of information.

References

[1] M. Milan Randić, On characterization of molecular branching, J. Am. Chem. Soc. 97 (1975) 6609–6615.
[2] L.B. Kier, L.H. Hall, W.J. Murray, M. Randić, Molecular connectivity. I. Relationship to nonspecific local anesthesia, J. Pharm. Sci. 64 (1975) 1971–1974.
[3] L.H. Hall, L.B. Kier, W.J. Murray, Molecular connectivity. II. Relationship to water solubility and boiling point, J. Pharm. Sci. 64 (1975) 1974–1977.
[4] L.B. Kier, L.H. Hall, Molecular connectivity. VII. Specific treatment of heteroatoms, J. Pharm. Sci. 65 (1976) 1806–1809.
[5] L.B. Kier, L.H. Hall, Molecular Connectivity in Chemistry and Drug Research, Academic Press, New York, 1976.
[6] F. Harary, Graph Theory, Addison-Wesley, Reading, MA, 1972.
[7] A.T. Balaban, Historical development of topological indices, in: J. Devillers, A.T. Balaban (Eds.), Topological and Related Indices, QSAR and QSPR, Gordon and Breach, London, 1999, pp. 21–58.
[8] H. Wiener, Structural determination of paraffin boiling points, J. Am. Chem. Soc. 69 (1947) 17–20.
[9] J.R. Platt, Prediction of isomeric differences in paraffin properties, J. Phys. Chem. 56 (1952) 151–174.
[10] H. Hosoya, A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons, Bull. Chem. Soc. Jpn. 44 (1971) 2332.
[11] I. Gutman, B. Ruscić, N. Trinajstić, C.F. Wilcox, Graph theory and molecular orbital XII acyclic polyenes, J. Phys. Chem. 62 (1971) 3399–3409.
[12] M. Randić, The connectivity index 25 years after, J. Mol. Graph Mol. Model 20 (2001) 19–35.
[13] D. Bonchev, Overall connectivity — a next generation molecular connectivity, J. Mol. Graph Mol. Model 20 (2001) 65–75.
[14] A. Balaban, New trends in molecular connectivity, J. Mol. Graph Mol. Model, 2001, in press.
[15] L.H. Hall, L.B. Kier, Development of molecular connectivity and the role of creativity, in: Abstracts of the Proceedings of the 220th National American Chemical Society Meeting, Washington, DC, 2000.
[16] D. Agin, L. Hersh, D. Holtzman, The action of anesthetics on excitable membranes: a quantum-mechanical analysis, Proc. Natl. Acad. Sci. U.S.A. 53 (1965) 952–958.
[17] L.B. Kier, Molecular Orbital Theory in Drug Research, Academic Press, New York, 1971, pp. 236–239.
[18] Ernest L. Eliel, Stereochemistry of Carbon Compounds, McGraw-Hill, New York, 1962, p. 1.
[19] L.B. Kier, L.H. Hall, Molecular Structure Description: The Electrotopological State, Academic Press, New York, 1999.
[20] L.B. Kier, L.H. Hall, Molecular Connectivity in Structure–Activity Analysis, Wiley, New York, 1986.
[21] J. Devillers, Neural Networks in QSAR and Drug Design, Academic Press, London, 1996.
[22] D. Bonchev, Information Theoretic Indices for Characterization of Chemical Structure, Wiley, Chichester, UK, 1983.
[23] R.B. King, Chemical Applications of Topology and Graph Theory, Elsevier, Amsterdam, 1983.
[24] A. Balaban, Applications of Graph Theory, Academic Press, London, 1976.
[25] J. Devillers, A. Balaban, in: J. Devillers, A.T. Balaban (Eds.), Topological and Related Indices in QSAR and QSPR, Gordon and Breach, London, 1999.
[26] S.C. Basak, B.D. Gute, G.D. Grunwald, A hierarchical approach to the development of QSAR models using topological, geometrical and quantum mechanical parameters, in: J. Devillers, A.T. Balaban (Eds.), Topological Indices and Related Descriptors in QSAR and QSPR, 1999, pp. 675–696.
[27] S.C. Basak, B.D. Gute, G.D. Grunwald, Use of topostructural, topochemical and geometric parameters in the prediction of vapor pressure: a hierarchical QSAR approach, J. Chem. Inf. Comput. Sci. 37 (1997) 651–655.
[28] S.C. Basak, B.D. Gute, G.D. Grunwald, Prediction of complement-inhibitory activity of benzamidines using topological and geometric parameters, J. Chem. Inf. Comput. Sci. 37 (1997) 651–655.
[29] L.B. Kier, L.H. Hall, General definition of valence delta-values for molecular connectivity, J. Pharm. Sci. 72 (1983) 1170–1173.
[30] L.B. Kier, L.H. Hall, Derivation and significance of valence molecular connectivity, J. Pharm. Sci. 70 (1981) 583–589.
[31] L.H. Hall, L.B. Kier, The molecular connectivity chi and kappa shape indices, in: D. Boyd, K. Lipkowitz (Eds.), Reviews in Computational Chemistry, VCH, New York, 1991, pp. 384–385.
[32] L.H. Hall, D. Stewart, Response surface analysis of bioconcentration by chlorinated organics using molecular connectivity, SAR QSAR Environ. Res. 2 (1994) 181–191.
[33] T. DiPaolo, L.B. Kier, L.H. Hall, Molecular connectivity and structure–activity relationship of general anesthetics, Mol. Pharm. 13 (1977) 31–37.

[34] G.L. Amidon, S.T. Anik, Comparison of several topological indexes with molecular surface area in aqueous solution estimation, J. Pharm. Sci. 65 (1976) 801–806.
[35] S.G. Gupta, P. Singh, The relationship of pi-binding with molecular connectivity in hydrocarbons, Bull. Chem. Jpn. 52 (1979) 2745–2750.
[36] V.K. Gombar, Molecular connectivity and it correlation with physicochemical properties. 1. Boiling points of saturated aliphatic carboxylic acids, Pharmacos 25 (1981) 46–51.
[37] L.B. Kier, L.H. Hall, R.J. Simons, Structure–activity studies on the mutagenicity of nitrosamines using molecular connectivity, J. Pharm. Sci. 67 (1978) 725–726.
[38] T.W. Schultz, L.B. Kier, L.H. Hall, Structure–toxicity relationships of selected nitrogenous heterocyclic compounds. III. Relations using molecular connectivity, Bull. Environ. Contam. Toxicol. 28 (1982) 373–378.
[39] L.H. Hall, L.B. Kier, Molecular connectivity of phenols and their toxicity to fish, Bull. Environ. Contam. Toxicol. 21 (1984) 354–362.
[40] L.B. Kier, L.H. Hall, Molecular connectivity study of muscarinic receptor affinity of acetylcholine antagonists, J. Pharm. Sci. 67 (1978) 1408–1412.
[41] L.H. Hall, E.L. Maynard, L.H. Kier, QSAR investigation of benzene toxicity to fathead minnow using molecular connectivity, Environ. Toxicol. Chem. 8 (1989) 783–788.
[42] S.C. Basak, G.J. Niemi, G.D. Veith, Predicting properties of molecules using graph invariants, J. Math. Chem. 7 (1991) 243–272.
[43] G.W. Adamson, J.A. Bush, Evaluation of an empirical structure–activity relationship for property prediction in a structurally diverse group of local anesthetics, J. Chem. Soc., Perkin Trans. 1 (1976) 168.
[44] G.W. Adamson, D. Bawden, An empirical method of structure–activity correlation for polysubstituted cyclic compounds using Wiswesser line notation, J. Chem. Inf. Comput. Sci. 16 (1976) 161.
[45] L.B. Kier, L.H. Hall, Intermolecular accessibility: the meaning of molecular connectivity, J. Chem. Inf. Comput. Sci. 40 (2000) 792–795.
[46] L.B. Kier, L.H. Hall, Molecular connectivity: intermolecular accessibility and encounter simulation, J. Mol. Graph Model 20 (2001) 76–83.
[47] M. Randić, Characterization of atoms, molecules, and classes of molecules based on path enumerations, MATCH 7 (1979) 3–60.
[48] L.H. Hall, L.B. Kier, Molecular connectivity chi indices for database analysis and structure–property modeling, in: J. Devillers, A.T. Balaban (Eds.), Topological and Related Indices in QSAR and QSPR, Gordon and Breach, London, 1999, pp. 307–360.
[49] S.C. Basak, V.R. Magnuson, G.J. Niemi, R.R. Regal, Determining structural similarity of chemicals using graph-theoretic indices, Discrete Appl. Math. 19 (1988) 17–44; Special Volume: J.W. Kennedy, L.V. Quintas (Eds.), Applications of Graph Theory in Chemistry and Physics.
[50] S.C. Basak, G.J. Niemi, G.D. Veith, Predicting properties of molecules using graph invariants, J. Math. Chem. 7 (1991) 243–272.
[51] D.E. Needham, I.-C. Wei, P.G. Seybold, Molecular modeling of the physical properties of the alkanes, J. Am. Chem. Soc. 110 (1988) 4186–4194.
[52] T. Zhu, P.G. Seybold, Explorations of molecular structure–property relationships, SAR QSAR Environ. Res. 10 (1999) 101–115.
[53] P.G. Seybold, M. May, U. Bagal, Molecular structure–property relationships, J. Chem. Educ. 64 (1981) 575–581.
[54] P. Seybold, M.A. May, M.L. Gargas, On the physical properties of halogenated hydrocarbons, Acta Pharm. Jugosl. 36 (1986) 253–260.
[55] M.L. Gargas, P.G. Seybold, M.E. Anderson, Modeling the tissue solubilities and metabolic rate constant (V_{max}) of halogenated methanes, Toxicol. Lett. 43 (1988) 235–256.
[56] A.T. Balaban, Using real numbers as vertex invariants for third generation topological indexes, J. Chem. Inf. Comput. Sci. 32 (1992) 23–28.
[57] A.T. Balaban, T.-S. Balagan, Correlations using topological indices based on real graph invariants, J. Chim. Phys. 89 (1992) 1735–1745.
[58] A.T. Balaban, Topological indices based on topological distances in molecular graphs, Pure Appl. Chem. 55 (1983) 199–206.
[59] D. Bonchev, N. Trinajstic, Information theory, distance matrix, and molecular branching, J. Chem. Phys. 67 (1977) 4517–4533.
[60] D. Bonchev, O. Mekenyan, N. Trinajstic, Isomer discrimination by topological information theory, J. Comput. Chem. 2 (1981) 127–148.
[61] D. Bonchev, Overall connectivity and topological complexity: a new tool for QSPR/QSAR, in: J. Devillers, A.T. Balaban (Eds.), Topological and Related Indices in QSAR and QSPR, Gordon and Breach, London, 1999, pp. 361–402.
[62] J. Galvez, R. Garcia-Domenech, J.V. de Julian-Ortiz, Topological approach to drug design, J. Chem. Inf. Comput. Sci. 35 (1995) 272–284.
[63] J. Galvez, R. Garcia-Domenech, J.V. de Julian-Ortiz, R. Soler, Topological approach to analgesia, J. Chem. Inf. Comput. Sci. 34 (1994) 1198–1203.
[64] R. Garcia-Domenech, F.J. Garcia-March, R.G. Soler, J. Galvez, G.M. Anton-Fos, J.V. de Julian-Ortiz, New analgesics designed by molecular topology, Quant. Struct. Act. Relat. 15 (1996) 201–207.
[65] J. Galvez, M.J. Gomez-Icehon, R. Garcia-Domenech, J.V. Castell, New cytostatic agents obtained by molecular topology, Bioorg. Med. Chem. Lett. 6 (1996) 2301–2306.
[66] L. Pogliani, Molecular connectivity model for determination of isoelectric point of amino acids, J. Pharm. Sci. 81 (1992) 334–336.
[67] L. Pogliani, Molecular connectivity: treatment of electronic structure of amino acids, J. Pharm. Sci. 81 (1992) 967–969.
[68] L. Pogliani, Molecular connectivity model for determination of physicochemical properties of α-amino acids, J. Phys. Chem. 97 (1993) 6731–6737.
[69] L. Pogliani, Modeling purines and pyrimidines with the linear combination of connectivity–molecular connectivity LCCI–MCI method, J. Chem. Inf. Comput. Sci. 36 (1996) 1082–1091.
[70] L. Pogliani, From molecular connectivity indices to semi-empirical connectivity terms: recent trends in graph theoretical descriptors, Chem. Rev. 100 (2000) 3827–3858.
[71] L.H. Hall, L.B. Kier, A molecular connectivity study of electron density in alkanes, Tetrahedron Lett. 33 (1977) 1953–1957.
[72] L.H. Hall, L.B. Kier, Molecular connectivity and substructure analysis, J. Pharm. Sci. 67 (1978) 1743–1747.
[73] L.H. Hall, L.B. Kier, Molecular connectivity of phenols and their toxicity to fish, Bull. Environ. Contam. Toxicol. 32 (1984) 354–362.
[74] Molconn-Z, a software package in FORTRAN available from Lowell H. Hall, Hall Associates Consulting, 2 Davis Street, Quincy, MA and from Glen Kellogg, Edusoft, LC, PO Box 1811, Ashland, VA.
[75] H.H. Maw, L.H. Hall, QSAR modeling of HIV-1 binding using only topological descriptors: the E-state model, J. Chem. Inf. Comput. Sci., submitted for publication.
[76] M.K. Holloway, J.M. Wai, T.A. Halgren, P.M.D. Fitzgerald, J.P. Vacc, B.D. Dorsey, R.B. Levin, W.J. Thompson, L.J. Chen, J. deSolms, N. Gaffin, G.K. Ghosh, E.A. Giuliani, S.L. Graham, J.P. Guare, R.W. Hungate, T. Lyle, W.M. Sanders, T.J. Tucker, M. Wiggins, C.M. Wiscount, O.W. Waltersdorf, S.D. Young, P.L. Darke, J.A. Zugay, A priori prediction of activity for HIV protease inhibitors employing for energy minimization in the active site, J. Med. Chem. 38 (1995) 305–317.
[77] J.D. Gough, L.H. Hall, Modeling the toxicity of amide herbicides using the electrotopological state, Environ. Toxicol. Chem. 18 (1999) 1069–1075.
[78] M. Parham, L.H. Hall, L.B. Kier, Accurate prediction of *n*-octanol/water partition coefficients using neural network algorithms and atom type E-state atom indices, in: Abstracts of the Proceedings of the 220th National American Chemical Society Meeting, Washington, DC, 2000.
[79] L.B. Kier, L.H. Hall, J.W. Frazer, Design of molecules from quantitative structure–activity relationship models. I. Information

transfer between path and vertex degree counts, J. Chem. Inf. Comput. Sci. 33 (1993) 143–147.
[80] L.B. Kier, L.H. Hall, J.W. Frazer, Design of molecules from quantitative structure–activity relationship models. II. Derivation and proof of information transfer relating equations, J. Chem. Inf. Comput. Sci. 33 (1993) 148–152.
[81] L.H. Hall, S. Robert, L.B. Kier, Design of molecules from quantitative structure–activity relationship models. III. Role of higher-order path counts: path three, J. Chem. Inf. Comput. Sci. 33 (1993) 598–603.
[82] L.B. Kier, L.H. Hall, The generation of molecular structures for a graph-based equation, Quant. Struct. Act. Relat. 12 (1994) 383–388.
[83] J. Fisk, L. Hall, Degree set generation for chemical graphs, J. Chem. Inf. Comput. Sci. 34 (1994) 1184–1189.