

---

## CHAPTER 9

# The Molecular Connectivity Chi Indexes and Kappa Shape Indexes in Structure–Property Modeling

Lowell H. Hall\* and Lemont B. Kier†

\**Department of Chemistry, Eastern Nazarene College, Quincy, Massachusetts 02170, and* †*Department of Medicinal Chemistry, School of Pharmacy, Virginia Commonwealth University, Richmond, Virginia 23298*

---

### INTRODUCTION

When a chemist contemplates the properties of a molecule, the usual beginning point is the structure of the molecule. What can be said about the magnitude of the property and its dependence on changes in the molecular structure depends on the chemist's ability to develop relations between structure and property. Often this relation is qualitative and referred to as SAR, a structure–activity relationship. In many physical–organic, biochemical, and biological areas, however, it is increasingly necessary to translate those qualitative ideas into quantitative relations expressed in algebraic equations known as QSAR, quantitative structure–activity relationships. For QSAR, then, molecular structure must be expressed in numerical form suitable for manipulation in algebraic equations. The insights into SAR and QSAR may come in most basic form from quantum chemical principles. For many practical problems in the design of new molecules, however, it is necessary to develop such form–function relationships in a different manner.

Chemical graph theory has been the basis for new applications in structure–property methods during recent years.<sup>1–10</sup> With an emphasis on the molecular skeleton, numerical indexes have been developed that represent features of molecular structure. These indexes are numerical values that parallel variation

in molecular structure and also parallel variation in properties. The most widely used of these indexes are the molecular connectivity indexes.<sup>2,3,8,9</sup> Applications have been made to solubilities,<sup>3,12</sup> boiling points,<sup>11,12</sup> densities,<sup>13</sup> molar refraction,<sup>2,3,11</sup> partition coefficients,<sup>2,14-16</sup> thermochemical properties,<sup>2</sup> and gas chromatographic retention parameters.<sup>3,15-24</sup> QSAR in biological studies have included anesthetic potency,<sup>25-29</sup> hallucinogenic activity,<sup>30-33</sup> enzyme inhibition,<sup>34,35</sup> flavor, odor and taste,<sup>36-39</sup> bioconcentration factors,<sup>40</sup> soil sorption,<sup>41</sup> antimicrobial and antibacterial activity,<sup>42-44</sup> toxicity,<sup>45-49</sup> carcinogenicity,<sup>50-53</sup> and others.<sup>2,3</sup> Reviews and monographs have appeared in the past several years.<sup>1,4,7-9</sup>

In the molecular connectivity method, numerical indexes are developed to represent and characterize molecular structure. The indexes are suitable for use in structure–property relationships that may be developed by standard regression methods, producing equations for estimation of property values. The indexes are also used as a basis for pattern recognition and discriminant analysis.<sup>54-58</sup>

These topological indexes, based on the molecular connectivity approach, include three types: the  ${}^m\chi_t$  molecular connectivity chi indexes that characterize the structural attributes of molecules,<sup>2,3</sup> the  ${}^m\kappa$  kappa indexes of molecular shape,<sup>59-61</sup> and the topological equivalence state  $T$  values that individually characterize atoms and groups in the molecular skeleton<sup>62</sup> and are used primarily to determine chemically equivalent atoms within a molecule. A further development of this approach has led to the electrotopological state atom indexes,<sup>63-66</sup> which will not be discussed here but will be presented elsewhere.<sup>67</sup> Molecular connectivity chi indexes are discussed in the first part of this paper along with illustrative applications. Then kappa shape indexes are discussed. The topological state index is discussed in the final section.

The term “molecular structure” is used in many ways; it sometimes appears interchangeably with other terms such as molecular geometry or architecture even though these terms are not synonymous. The term structure refers to the constituent atoms and their relationships, whereas the term geometry refers to specific three-dimensional arrangements. For a definition of molecular structure, we have adopted the statement given by Eliel,<sup>68</sup> as follows, “The structure of a molecule is completely defined by the number and kind of atoms and the linkages between them.” The term molecule “topography” is applied to three-dimensional structure, whereas molecular “topology,” the set of atoms and connections, refers to the molecular structure as defined by Eliel. Whatever the consequences of each of these molecular aspects, the three-dimensional structure follows from the molecular topology. In our discussions of chemical graph theory, the focus of attention is on ways of characterizing the topology of molecules using numerical indexes and applying these indexes to investigation in QSAR.

Because of the importance of structure–property models, it has become increasingly important that broadly based methods be developed for the establishment and investigation of such relations. It is both timely and natural that the methods of graph theory be applied to structure–property modeling. The

chemical graph is a versatile vehicle for the rich expression of chemical structure information. The developments of the past decade are a clear indication of the fertility of this area of investigation.

## BACKGROUND FOR MOLECULAR CONNECTIVITY

The first step in graph theoretical analysis of molecular structure is adoption of an appropriate structure representation. Starting with the definition of molecular structure, attention is centered on the molecular skeleton, which consists of the network of chemical bonds, including the set of atoms and connections between atoms. Such an entity may be described by the molecular graph, which consists of vertexes, representing skeletal atoms, and edges, representing skeletal bonds.<sup>69-72</sup> The term bond is not necessarily a covalent bond between a pair of adjacent atoms because the vertex can be a single atom or an atom with its bonded hydrogen atoms, a skeletal hydride group such as  $-\text{CH}_2-$  or  $-\text{NH}_2$  or  $-\text{OH}$ . The "connection" may be a given covalent bond or it could include another set of interactions.

The approach used in chemical graph theory is to abstract from the molecular structure those elements that lead to structure variables in the form of numerical indexes. The set of atoms and connections is viewed as structure information but in a form not amenable directly to QSAR analysis. The first step is to adopt a form for the molecular skeleton as the basis for extraction of structure information. To represent the molecular skeleton, the hydrogen-suppressed graph is most commonly used<sup>1-4</sup>; hydrogen atoms are not explicitly considered; hydrogen atoms are incorporated in skeletal groups which are the graph vertexes.

Figure 1 shows examples of the relation between the usual form of molecular structural formula and the hydrogen-suppressed graph. Carbon atom symbols are usually omitted from the graph, but heteroatoms are usually written explicitly. It may be necessary to include hydrogen atoms that are part of functional groups to distinguish among different fragments. These hydrogen atoms, however, serve only to distinguish; they are not part of the set of vertexes and edges. In performing calculations it is assumed that any vertex may also contain sufficient hydrogen atoms to satisfy the carbon valence.

The lowest level structure descriptor is simply the count of atoms. However, only in alkanes are the properties even roughly proportional to the number of atoms. More information-rich structure indicators are needed. In the earliest topological methods, numerical indexes were developed simply by counting the number of edges or pairs of adjacent edges. Wiener,<sup>73,74</sup> Platt,<sup>75,76</sup> and others<sup>77-84</sup> developed structure indexes that produced interesting correlations to some properties of saturated hydrocarbons. As impressive as this work is, it could not be readily or usefully extended to unsaturated compounds or to heteroatom-

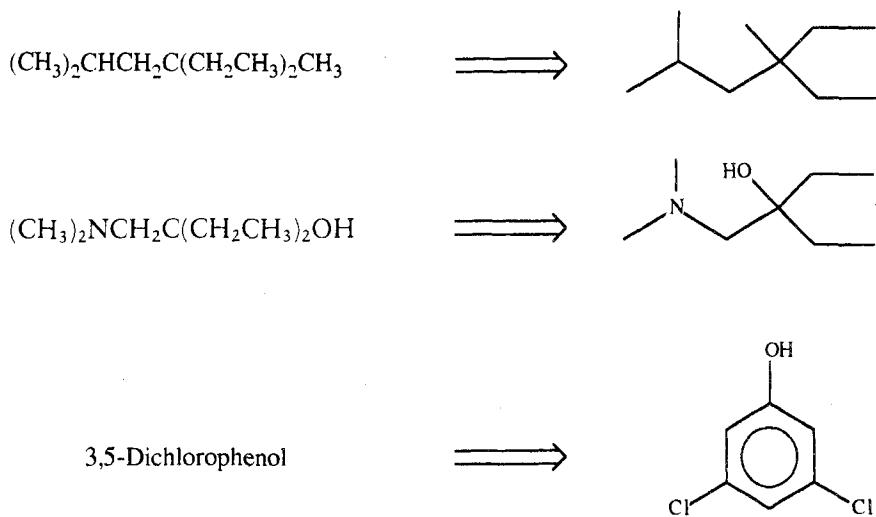


Figure 1 The relationship between molecular formulas and the corresponding hydrogen-suppressed graphs.

containing molecules at all. Counting edges, while abstracting some important information, ignores most of the important information contained in the molecular skeleton.

Simple edge counting gives equal weight to  $\text{CH}_3-\text{CH}$ ,  $\text{CH}_2-\text{CH}_2$ ,  $\text{CH}-\text{C}$ ,  $\text{CH}_2=\text{CH}$ , etc. Even worse, simple edge counting equates edges such as  $\text{CH}_2-\text{CH}_2$ ,  $\text{CH}-\text{NH}_2$ ,  $\text{CH}_2-\text{Cl}$ ,  $\text{CH}_2-\text{OH}$ , etc. The earlier methods gave equal weights to chemically nonequivalent groups. A direct approach to solving this problem is the use of appropriate edge weights. Randić was the first to propose the use of the vertex degree<sup>10</sup> in edge weights. Vertex degree,  $v$ , for a given vertex is defined as the number of neighboring vertexes. A methyl group possesses only one neighbor,  $v = 1$ ; a methylene group has two,  $v = 2$ ; and a hydroxyl group has one,  $v = 1$ . Randić also proposed an edge weight as  $(v_i v_j)^{-1/2}$  for the edge between vertexes  $i$  and  $j$ . Randić defined a branching index as the sum of edge weights for all edges in a molecular graph:  $\sum(v_i v_j)^{-1/2}$ . This development is appropriate only for saturated alkanes<sup>2,3,4,10</sup> because there is no provision for atom identity, bond types, or number of hydrogen atoms in each skeletal group (except for the case of saturated hydrocarbons.) This branching index is essentially a mathematical property of graphs; it does not adequately represent molecular graphs. Further, a single index would appear to be insufficient to relate to the wide variety of molecular properties, especially when biological and environmental properties are of interest.

---

## DEVELOPMENT OF MOLECULAR CONNECTIVITY

The objective of the molecular connectivity method is the quantitation of molecular structure based on the topological and electronic character of the atoms in the molecule. The pursuit of this objective leads to topological indexes derived from graph theoretic concepts, definitions, and procedures. We need to have methods for broad application in structure-property models, especially QSAR models. From these methods, we seek a basis for better understanding of the relation between structure and properties.

Topological indexes should be based on the most significant features of molecular structure, including atom identity, bonding environment of each atom, and the existence of hydrogen atoms bonded to the skeletal atom. Each of these aspects is distinctly electronic in nature.

The identifying characteristics of atoms include atomic number and number of electrons partitioned between valence electrons and core electrons. The immediate bonding environment of atoms in the molecular skeleton depends on the number and arrangement of the valence electrons and the number and type of bonds. In most graph theoretical methods, a hydrogen-suppressed skeleton is used to facilitate the counting and enumeration of skeletal features.

It seems clear that the electronic structure of the bonded atoms must be encoded in the topological indexes if they are to parallel variation in molecular properties. Otherwise, the resulting numerical quantities represent only the mathematical properties of graphs, rather than molecules, which are the chemical entities of interest.

Beyond the consideration of important atomic information to be encoded, we must examine the nature of the properties of interest and their dependence on structure. There is rich information in the various relations between the structure of molecules and their properties. Properties depend on molecular structure in many different ways. Thus, topological indexes must be developed to deal with all the different cases.

Properties relate to measures of molecular size in various ways. Some properties depend heavily on the atom count. Molar volume, heat of atomization, and molar refraction are approximately proportional to atom count; contributions can be summed to give an estimate of the total property value. Generally, properties require additional information about the immediate bonding environment of atoms for useful estimation models. Water solubility depends on the branching in the molecular skeleton. Molecular fragments larger than individual atoms are required to estimate most property values. This additive-constitutive nature of properties has been amply demonstrated by many additive property schemes.<sup>84-98</sup>

For certain classes of molecules and properties, additivity schemes work well. However, these methods are not structure-based in the sense used here. Naming a fragment and assigning a property value does not represent the structure of the molecule. The need for structure-based approaches becomes especially evident when one wishes to invert the usual structure–property relation to predict a structure for a molecule that possesses a predetermined property value.

Some properties do depend heavily on atom count but others depend not so much on the structure of the overall molecule as on the specific bonding in a local skeletal area or even on a single atom. Ionization potential of monosubstituted alkanes is one example of such structure–property dependence. The Woodward rules for estimation of ultraviolet absorption in organic molecules is another example of such a structure–property scheme.<sup>99</sup>

In biological properties, the structure dependence is much less on atom count than on larger structure fragments. The nature of a pharmacophore indicates that a specific arrangement of certain atoms is required for activity, whereas variation of another molecular region may account for variation in activity. Further, the structural relationship between atoms that are not directly bonded is often important in the variation of biological activity.

Some properties such as overall molecular shape or compactness may depend on the entire topological framework or skeleton. The symmetry or topological equivalence properties of atoms in molecules also depend on the whole molecular structure. Some aspects of symmetry and shape will be discussed later.

A general examination of structure–property relations clearly indicates that there is much information required to describe the structure of a molecule, particularly when structure–property relations are to be developed. Some of the pieces of information, which may be developed from characterization of molecular structures, may be interrelated while others may be independent. As an extreme case, in a set of normal alkanes the number of carbon atoms, the number of hydrogen atoms, the number of methylene groups, and the number of carbon atoms bonded to two other carbon atoms are all exactly correlated. When a set of normal and branched alkanes is considered, this exact correlation is broken up, but there remain interrelations among the quantities. Other quantities such as the number of phenyl rings and the number of carbonyl groups in a given set of molecules may be rather unrelated. To put it another way, the various structure features of molecules are not usually unrelated (orthogonal in mathematical terms). If orthogonality is important in a given investigation, then that orthogonality must be introduced.

When a wide range of structure–property relations is examined, it is clear that many topological indexes are required to express the wide range of structure characteristics of covalent molecules. Although it is useful and interesting to develop a unique and highly discriminating index, such a single index is insufficient in itself for the broad range of QSAR studies.

A final consideration in QSAR relates to the nature in which structure information is expressed. It might be said that there are two aspects to the informa-

tion, the structure information itself and the structure complexity or diversity.<sup>100</sup> Although certain properties may in some way depend on a measure of the amount of information, most properties certainly depend on the information itself. Most properties of interest depend much more on atom type, atom count, and the structure relationships in the skeleton than on the amount of information represented by the structure. Certainly properties, such as those related to symmetry, complexity, or entropy, do have a heavy dependence on the amount of structure information. Measures of the amount of information are therefore necessary but incomplete in the application of graph theory to structure-property relationships.

---

## MOLECULAR CONNECTIVITY APPROACH

With these considerations as a general background, let us summarize our approach. In the molecular connectivity approach, the molecule is represented by the hydrogen-suppressed graph. The key feature in the quantitation of the graph is the characterization of the atom in the molecular skeleton. The molecular connectivity method explicitly introduces the electronic character of atoms into the graphic representation of molecules. Atom identity is specified through the molecular connectivity delta values: the simple delta,  $\delta$ , and the valence delta,  $\delta^v$ .<sup>101,102</sup>

The simple delta value for a skeletal atom,  $\delta$ , is the number of skeletal neighbors, which is equivalent to the vertex degree, as used in graph theory. However, Kier and Hall have pointed out that this defined graph theory quantity, "number of skeletal neighbors," has a direct electronic interpretation.<sup>2,3</sup> The number of skeletal bonds to a vertex (skeletal atom) is equal to the number of electrons (from that skeletal atom) assigned to sigma type orbitals less the number of electrons assigned to sigma bonds to hydrogen atoms, as expressed in Eq. [2].<sup>3,102</sup>

$$\delta_i = \text{number of skeletal neighbors of atom } i \quad [1]$$

also

$$\delta_i = \sigma_i - b_i \quad [2]$$

where

$$\sigma_i = \text{number of atom } i\text{'s electrons in sigma orbitals} \quad [3]$$

and

$$b_i = \text{number of hydrogen atoms bonded to skeletal atom } i \quad [4]$$

Accordingly, this simple graph quantity is conferred with chemical electronic meaning, and chemical graph theory is born. To increase significantly the amount of electronic information, Kier and Hall developed the valence delta value for explicit encoding of atom identity along with bonding environment and number of bonded hydrogens.<sup>2,3,101,102</sup> For atom identity all the valence electrons must be counted, not just those involved in skeletal bonding. For first-row atoms in covalent molecules, the valence delta for atom  $i$ ,  $\delta_i^v$ , is

$$\delta_i^v = Z_i^v - h_i \quad [5]$$

where  $Z_i^v$  = the number of valence electrons for skeletal atom  $i$ . Atoms could be identified by their atomic symbols, but such identification is a name or identification tag that bears no direct relation to electronic properties. In the valence delta value, the atom is identified implicitly along with its immediate bonding environment. A carbon atom involved in a single bond,  $\text{CH}_3-$ , or in a double bond, such as  $-\text{CH}=$ , has a different  $\delta^v$  than a triple bonded carbon,  $-\text{C}\equiv$ , simply on the basis of a different value for  $h_i$ . The number of hydrogen atoms in the skeletal group is also explicitly included as  $h_i$ , which distinguishes  $-\text{CH}_3$  from  $-\text{NH}_2$  from  $-\text{OH}$  from  $-\text{F}$ . In this way, although chemical graph theory uses the hydrogen-suppressed graph, the hydrogen atom count is included by the use of the delta values.

For atoms beyond fluorine, the principal difference among family members is the number of core electrons. The valence delta value must explicitly take this factor,  $Z - Z^v$ , into account. ( $Z$  is the atomic number.)

$$\delta^v = (Z^v - h)/(Z - Z^v - 1) \quad [6]$$

Thus, for a given atom the valence delta takes into account the number of valence electrons and the number of core electrons.<sup>3,102</sup>

This pair of delta values is seen as a characterization of the atom in its valence state. The simple delta,  $\delta$ , describes the role of the atom in the skeleton in terms of its connectedness and count of sigma electrons; it could be called the sigma electron descriptor. The valence delta,  $\delta^v$ , encodes the electronic identity of the atom in terms of both valence electron count and core electron count. It could be called the valence electron descriptor. The isolated, unbonded atom may be thought of as characterized by its atomic number,  $Z$ , and the number of valence electrons,  $Z^v$ . In its valence state, the bonded atom is characterized by  $\delta$  and  $\delta^v$ . Embedded in the molecular skeleton, the full characterization of the atom in the environment of the whole molecule is given by the topological equivalence value, described in a later section, and the electrotopological state value, presented separately.<sup>67</sup> A representation of the whole molecule is accomplished by the combination of chi, kappa, and topological state indexes.

The pair of delta values is useful in their own right, in addition to their use in chi indexes. Kier and Hall demonstrated the relationship between these delta

values and two important valence state properties. First, the sum of the delta values correlates highly with the van der Waals volume for skeletal groups in their valence states,<sup>101</sup> using atomic volume values developed by Bondi.<sup>103</sup>

$$V_{vdw} = 17.03 - 1.59(\delta^v + \delta) \quad [7]$$

$$r = 0.990, s = 0.52, F = 8.59, n = 20$$

Further, in the same paper, Kier and Hall demonstrated for the first time a simple and direct relation between electron counts and valence state electronegativity.<sup>101</sup>

$$X_{MJ} = 7.99(\delta^v - \delta)/N^2 - 7.07 \quad [8]$$

$$r = 0.988, s = 0.48, F = 660, n = 19$$

where  $X_{MJ}$  is the Mulliken valence state electronegativity values developed by Jaffé et al.<sup>104,105</sup> and  $N$  is the principal quantum number. This work led to the definition of the Kier–Hall relative electronegativity value as  $(\delta^v - \delta)/N^2$ . This value is relative to the electronegativity of carbon sp<sup>3</sup> taken as zero. (Note that in this usage for electronegativity,  $\delta^v = Z^v - h$  for all atoms because the  $N^2$  factor accounts for the varying distance of valence electrons from the nucleus and the count of core electrons is not used as it was in Eq. [8].) In a subsequent publication, Kier and Hall investigated the relation between relative electronegativity and electronic substituent parameters.<sup>106</sup> A model was developed as a basis for estimation of the Hammett sigma parameters for benzene substituents.

---

## MOLECULAR CONNECTIVITY METHOD

As these ideas were developed, it appeared appropriate to give a name to this approach. It was decided to use a name that brings together graph notions, such as adjacency and connectedness, with the central focus of this method, the chemical and molecular nature of structure. The name molecular connectivity was selected.<sup>2</sup>

The pair of connectivity delta values satisfied the first consideration stated above, the encoding of atom electronic character. It remained to develop a means of expressing the wide variety of structure information necessary to reveal adequately structure–property relations. The Randić edge weight expression appeared to be a useful way to encode the atom characteristics, but the Randić limitation to single edges prevented the full expression of the molecular structure information.<sup>2</sup>

In the molecular connectivity method a series of indexes is developed, expressing structure information at several levels. The molecular skeleton is conceived as consisting of fragments of different sizes and complexity. The molecular graph may be decomposed into fragments called subgraphs, such as a skeletal bond, a pair of adjacent bonds, a cluster of bonds to a central atom, etc.

This concept of graph decomposition into subgraphs (or molecular fragments) is illustrated in Figure 2 for the simple molecule isopentane. The various orders and types are shown for each of the distinct subgraphs in the molecule. The actual subgraph is shown in bold superimposed on the whole molecule graph in dashed lines. Some types of subgraphs are not found in isopentane, such as the cluster-4 and ring fragments, including the chain-3 and chain-4, all of which are shown only in the light line form.

A chi index is a weighted count of a given type of subgraph. There are two attributes of each index, the order and the type. The order,  $m$ , of a chi index is the number of graph edges in the corresponding subgraph. The type,  $t$ , refers to the particular arrangement of the edges in the subgraph such as Path (P), Cluster (C), Path/Cluster (PC), and Ring (CH).

### Order Zero: ${}^0\chi$

The simplest subdivision of a molecular skeleton or graph is the set of skeletal groups or vertexes. Because an individual vertex possesses no edges, the order (number of edges) is zero. The number of subgraphs of order zero is simply the number of skeletal atoms or vertexes,  $A$ . Each skeletal atom is characterized by its simple and valence delta values. The weight assigned to each vertex is the reciprocal square root of the delta value. The subgraph connectivity term,  $c$ , is defined for the zeroth order for both the simple and valence forms for atom  $s$ :

$${}^0c_s = (\delta_s)^{-1/2} \quad \text{and} \quad {}^0c_s^v = (\delta_s^v)^{-1/2} \quad [9]$$

The corresponding chi index of zero order is the sum of these terms for all atoms in the graph. Both a simple index, based on  $\delta_i$ , and a valence index, based on  $\delta_i^v$ , are defined.

$${}^0\chi = \sum {}^0c_s \quad \text{and} \quad {}^0\chi^v = \sum {}^0c_s^v \quad [10]$$

The summation is over the  $A$  atoms of the skeleton. The zero order chi index carries a low level of structure information. Little of the connectedness of the skeletal network is encoded; only the fact of the presence of the nearest neighbor to each atom is encoded. In the  ${}^0\chi^v$  index, atom identities are quantitated.

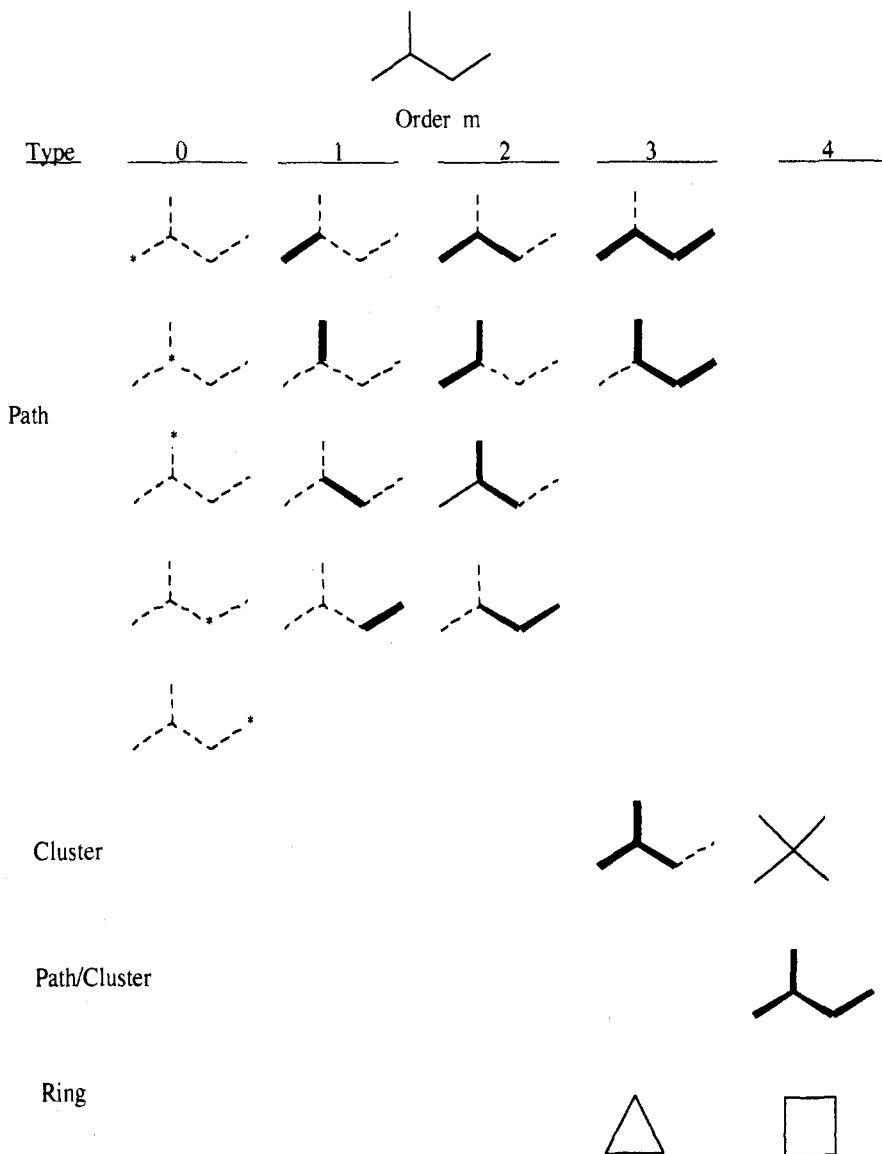


Figure 2 The isopentane hydrogen-suppressed graph and the various types of subgraphs that can be obtained from the graph. Subgraphs are shown in bold lines with the remainder of the skeleton shown in dashed lines. Shown in light lines are two ring-type subgraphs ( $t = \text{CH}$ ) and one cluster ( $m = 4$ ), which are not possible in the isopentane molecule.

**Order One:  ${}^1\chi$** 

The next higher order subgraphs are the graph edges. Subgraph terms  ${}^1c_s$  can be defined in analogy to those in Eq. [9] for the  $s$ th edge, that is, the edge between vertexes  $i$  and  $j$ :

$${}^1c_s = (\delta_i \delta_j)_s^{-1/2} \quad \text{and} \quad {}^1c_s^v = (\delta_i^v \delta_j^v)_s^{-1/2} \quad [11]$$

The first-order chi index is defined similarly to  ${}^0\chi$  with the summation over all the graph edges.

$${}^1\chi = \sum {}^1c_s \quad \text{and} \quad {}^1\chi^v = \sum {}^1c_s^v \quad [12]$$

The number of graph edges varies with type of structure. For acyclic molecules the number of skeletal bonds is one less than the number of skeletal atoms. The number of skeletal bonds is also referred to as the number of paths of length one,  ${}^1P$ . So,

$${}^1P = A - 1 \quad [13]$$

for acyclic molecules. For cyclic molecules there is a general relation between the number of edges,  ${}^1P$ , the number of skeletal atoms,  $A$ , and the number of rings,  $R$ :

$$R = {}^1P + 1 - A \quad [14]$$

The summation in Eq. [12] is over the number of edges,  ${}^1P$ . The quantity  $R$  is also known as the cyclomatic number of the graph.

The first-order chi indexes contain more structure information than do the zero-order indexes. The immediate bonding environment of each atom is encoded by virtue of the edge weight. Further, the number of terms in the sum,  ${}^1P$ , is dependent on the graph type, especially on the number of cycles or rings. The  ${}^1\chi^v$  index encodes both the atom identities as well as the connectedness in the molecular skeleton.

**Higher Order Chi Indexes:  ${}^m\chi_t$  and  ${}^m\chi_t^v$** 

For the order  $m \geq 2$  there are several types of subgraphs as illustrated in Figure 2: Path, Cluster, Path/Cluster, and Chain. For the third order, however, there is no Path/Cluster type. Each type of index is defined in a manner analogous to the lower order indexes as a product of reciprocal square roots.

$${}^m c_s = \Pi(\delta_i)_s^{-1/2} \quad \text{and} \quad {}^m c_s^v = \Pi(\delta_i^v)_s^{-1/2} \quad [15]$$

Graph	${}^0p$	${}^1p$	${}^2p$	${}^3p$	${}^4p$	${}^5p$	${}^0\chi$	${}^1\chi$	${}^2\chi$	${}^3\chi_p$	${}^4\chi_p$	${}^5\chi_p$
	6	5	4	3	2	1	4.828	2.914	1.707	0.957	0.500	0.250
	6	5	5	4	1	0	4.992	2.808	1.922	1.394	0.289	0.000
	6	5	5	3	2	0	4.992	2.770	2.183	0.866	0.577	0.000
	6	5	6	4	0	0	5.155	2.643	2.488	1.333	0.000	0.000
	6	5	7	3	0	0	5.207	2.561	2.914	1.061	0.000	0.000
	6	5	4	3	2	1	4.276	2.952	1.431	0.762	0.362	0.112
	6	5	5	4	1	0	4.439	2.489	1.470	0.943	0.289	0.000
	6	5	7	3	0	0	4.654	2.284	2.916	0.865	0.000	0.000
	7	7	8	8	8	8	3.834	2.134	1.336	0.756	0.428	0.029

Figure 3 The set of hexane isomers, three pentanol isomers, and phenol with the path counts and chi indexes, showing how these counts and indexes vary with structure.

The product is over the  $m + 1$  delta values in the subgraph of order  $m$ , except for the chain (or ring) type in which  $m$  is the number of atoms in the ring.

$${}^m\chi_t = \sum {}^m c_s \quad \text{and} \quad {}^m\chi_v = \sum {}^m c_s v \quad [16]$$

The summation is over the total number of subgraphs of order  $m$  and type  $t$  in the graph. Some examples of subgraph counts are given in Figure 3. The second-order count of paths,  ${}^2P$ , and higher orders, unlike the lower order subgraphs, are quite dependent upon the structure of the molecular skeleton. This effect is illustrated in Figure 3 for the isomers of hexane and some related structures. For the first three lower order indexes ( $m < 3$ ), there is only one type of subgraph. For this reason, the first three chi indexes are not given a type designation.

The chi indexes, especially those of higher order, exhibit both additive and constitutive character. Because they are by definition summations, the chi indexes are additive. Also the number of terms in each index summation depends in detail on the molecular structure described by the graph. As shown in Figure 3, the number of contributing subgraphs varies with structure, reflecting the constitutive nature of structure, and, hence, of the chi indexes. Further,

the very composition of the subgraph term as a product of delta values found in each subgraph also encodes the constitutive nature of the graph.

The chi indexes represent the whole graph. Each is composed of terms from the various parts of the graph, but the summation covers the whole graph. In this manner each index encodes in a particular manner the structure information resident in the molecular skeleton.

It is also possible to use chemical graph theory to characterize molecular shape and the individual skeletal atoms. However, before those descriptions are developed, several applications of molecular connectivity chi indexes will be presented.

## QSAR APPLICATIONS OF MOLECULAR CONNECTIVITY CHI INDEXES

A few examples will illustrate the wide applicability of the chi indexes in QSAR. Several references give many examples, some with a wide variety of studies,<sup>2-4</sup> some with physicochemical properties as by Seybold,<sup>107,108</sup> and some with properties of biological and environmental interest by Sabljić and Trinajstić<sup>9</sup> and Rouvray.<sup>109</sup>

### Chromatographic Retention

Investigation of chromatographic retention is one of the most active areas for QSAR studies using connectivity indexes and other topological indexes. Many papers have been written for this important area of analytical chemistry.<sup>3</sup> Sabljić demonstrated that the connectivity indexes are very useful in the QSAR analysis of chromatographic retention<sup>110</sup> and lists several references. In particular, he refers to analysis of chlorinated alkanes and benzenes. For 13 chlorinated benzenes he showed that the first-order index gives excellent correlation ( $r = 0.997$  and 0.985) for retention on SE-30 and on Carbowax 20M, respectively. These regressions were better than ones based on the use of the Wiener number or the Balaban  $J$  index, even with a heteroatom modification. The isomer pairs 1,3- and 1,4-dichlorobenzene as well as the other isomers are not discriminated by the first-order index alone.

Sabljić went on to show<sup>110</sup> that addition of the  ${}^4\chi_{PC}$  index provides the proper discrimination and improves the statistics to the experimental level of error, as follows:

$$I^{SE-30} = 337.0 {}^1\chi + 47.9 {}^4\chi_{PC} - 329.9 \quad [17]$$

$$r = 0.998, F = 1051, n = 13$$

This equation is statistically significant at the 99.9% level, and the indexes are individually significant at the 99.95 and 95% level, respectively. The  ${}^1\chi$  index gives good account of the dependence of  $I^{SE-30}$  on size, and the  ${}^4\chi_{PC}$  index provides proper isomer discrimination.

For the polar stationary phase, higher order chi indexes are required to provide the high quality correlation:

$$I^{CW-20M} = 226.8 {}^3\chi_P + 1588.0 {}^7\chi_{CH} + 649.1 \quad [18]$$

$$r = 0.998, F = 1347, n = 13$$

This equation is statistically significant at the 99.9% level; the indexes are significant at the 99.95% level. The size of the benzenes is encoded in the  ${}^3\chi_P$  index, whereas the isomer discrimination is in the  ${}^7\chi_{CH}$  index, which is based on toluene-like skeletal fragments.

Lee and White developed excellent data on polycyclic aromatic hydrocarbons, from naphthalene up to dibenzo[def, mno]chrysene.<sup>111</sup> The single index  ${}^1\chi$  yielded a QSAR with a standard deviation  $s = 5.8$ . The following three-variable equation gives an improved relation<sup>3</sup>:

$$RI = 35.0 {}^0\chi^v + 9.20 {}^4\chi_P + 100.0 {}^6\chi_{CH} - 38.0 \quad [19]$$

$$r = 0.999, s = 3.85, F = 5021, n = 32$$

Kaliszan and Foks investigated the use of the simple first-order connectivity index for the  $R_M$  values for reversed-phase thin-layer chromatographic data.<sup>112</sup> A data set consisting of 20 ring-substituted pyrazine carbothioamides, with nitrogen and oxygen-containing substituents, led to the following equation:

$$R_M = 0.733 {}^1\chi - 0.225 \quad [20]$$

They only reported the correlation coefficient,  $r = 0.946$ . Further, the chi-one index was calculated only for the substituent, a useful shortcut to quick calculation. Subsequently, Hall<sup>113</sup> improved the analysis by computing the chi indexes for the whole molecule and by using valence indexes.

$$R_M = 0.402 {}^1\chi - 0.225 {}^3\chi_{CV} - 1.994 \quad [21]$$

$$r = 0.973, s = 0.19, F = 148, n = 20$$

No residual exceeds two standard deviations; both aromatic and nonaromatic substituents are included. This two variable equation gives an excellent account of the data.

A group of 32 alcohols was investigated by Spivakovskii et al.<sup>114</sup> using a group additive scheme; they reported a 14-variable relation for a QSAR equation.

Because chi indexes encode structure in a more economical manner, a statistically improved model can be developed using fewer variables.<sup>3</sup>

$$RI = 1010 {}^1\chi - 770.7 {}^1\chi^v + 123.2 {}^2\chi - 136.7 {}^2\chi^v - 37.98 {}^4\chi_{PC} - 62.47 {}^5\chi_{PC}^v - 261.1 \quad r = 0.999, s = 5.78, F = 1647, n = 32 \quad [22]$$

The data set includes pentanols through nonanols in various isomeric forms. No residual exceeds two standard deviations. It should be noted that two pairs of simple and valence chi indexes appear in the equation. A 22% improvement in the standard deviation can be achieved by including the sums and differences in the set of variables from which the regression equation is constructed. The sum-variables are found to be most important, as follows:

$$RI = 403.4 {}^1\chi - 127.6 {}^0S + 88.5 {}^2S + 89.8 {}^3\chi_{PC}^v - 51.0 {}^5\chi_{PC}^v - 35.5 {}^3\chi_C + 373.9 \quad r = 0.9992, s = 4.5, F = 2707, n = 32 \quad [23]$$

The average relative error  $\epsilon = 0.70\%$ . The use of sum and difference variables usually permits either a better correlation for the same number of regression variables or fewer variables to accomplish an equivalent quality of correlation.

Recent work by Robbat et al.<sup>115-117</sup> has shown the utility of chi indexes in both the prediction of retention indexes and the identification of compounds from the retention index for polycyclic aromatic hydrocarbons and nitrated materials.

## Molar Volume

Molar volume can be measured by determination of density of dilute solutions with extrapolation to infinite dilution or by estimation, using the neat liquid density. Edward<sup>118</sup> has presented excellent data for alkanes. Hall and Kier have shown that these data,<sup>3</sup> plus six compounds added by Edward (taken from Longworth), lead to the following QSAR relation:

$$V_m(\text{cm}^3/\text{mol}) = 24.87 {}^1\chi + 11.86 {}^2\chi - 2.844 {}^4\chi_{PC} + 39.79 \quad r = 0.9999, s = 1.17, F = 86615, n = 37 \quad [24]$$

These three chi indexes are not highly intercorrelated. The  ${}^1\chi$  index encodes size and branching information. The  ${}^2\chi$  index encodes even more specific information about skeletal branching.  ${}^2\chi$  increases with the increase in skeletal branching; in the hexane series the  ${}^2\chi$  values increase in the order hexane < 3-methylpentane, < 2-methylpentane, < 2,3-dimethylbutane, < 2,2-dimethylbutane. The  ${}^4\chi_{PC}$  index is sensitive to specific structural aspects, especially to *gem* and *vicinal* substitution patterns. The negative sign on  ${}^4\chi_{PC}$  reflects a

volume decrease probably arising from structural accommodation for sterically crowded skeletons.

Seybold et al.<sup>107</sup> examined properties of chlorinated hydrocarbons. Included in that study is the QSAR equation for molecular volume:

$$V_m(\text{cm}^3/\text{mol}) = 11.9 {}^0\chi^v + 3.42 {}^3\chi_{\text{Cl}}^v - 9.65 \chi_{\text{Cl}}^v + 39.89 \quad [25]$$

$$r=0.9839, s=3.14, p<0.0001, n=24$$

The  $\chi_s^v$  is the valence structure index introduced by Sabljić<sup>24</sup> and is defined as

$$\chi_s^v = \sum \Pi (\delta_i^v)^{-1/2} \quad [26]$$

This index is a representation of the whole molecule and has been shown to be useful by Sabljić.<sup>118</sup>

### Heat of Atomization of Hydrocarbons and Alcohols

Heat of atomization has been carefully studied experimentally and has been shown to be an additive property.<sup>119</sup> For alkanes, using standard multiple linear regression statistics, it is possible to obtain a QSAR equation which possesses very good statistics.<sup>2</sup>

$$\Delta H_a = 286.38 A - 12.46 {}^1\chi + 1.515 {}^4\chi_{\text{Pc}} + 1.142 {}^4\chi_{\text{PC}} - 2.474 {}^5\chi_{\text{Cl}} - 2.026 {}^5\chi_{\text{PC}} + 114.38 \quad [27]$$

$$r>0.99999, s=0.46$$

where  $A$  is the number of carbon atoms,  $r$  is the correlation coefficient, and  $s$  is the standard deviation of the regression.

The chi indexes may be viewed as a basic set of variables from which useful combinations may be developed to emphasize certain structure features. With such transformed chi indexes, it is also possible to set up QSAR equations, often with fewer variables, which also permit structure comparisons among different classes of molecules. To assist in that comparison, we introduce a transformed set of chi indexes. The delta chi index is the difference between a simple index and the corresponding valence index. In this example,  $\Delta^2\chi$  is used and defined as follows:

$$\Delta^2\chi = {}^2\chi - {}^2\chi^v \quad [28]$$

The simple index encodes information only about the sigma electron network, whereas the valence index also encodes information about the nonsigma electrons. By virtue of taking the difference, the delta chi index emphasizes the contribution of the nonsigma electrons. For the alkenes and alkylbenzenes,  $\Delta^2\chi$

specifically encodes structure information about the electrons in  $\pi$  orbitals. In the alcohols,  $\Delta^2\chi$  encodes information about lone pair electrons on the hydroxyl group. The following cases summarize the heat of atomization results for four molecular classes:

Alkanes:

$$\Delta H_a = 279.60 A + 2.01 \Delta^2\chi - 0.190 \chi_{PC}^4 + 116.79 \quad [29]$$

$r > 0.99999, s = 0.83, \varepsilon = 0.032\%, F = 3477959, n = 44$

Alkenes:

$$\Delta H_a = 278.00 A + 15.33 \Delta^2\chi + 1.12 \chi_{PC}^4 - 22.61 \quad [30]$$

$r > 0.99999, s = 1.18, \varepsilon = 0.063\%, F = 1559107, n = 37$

Alkylbenzenes:

$$\Delta H_a = 235.53 A + 15.91 \Delta^2\chi + 1.56 \chi_{PC}^4 - 377.01 \quad [31]$$

$r > 0.99999, s = 0.31, \varepsilon = 0.015\%, F = 12378965, n = 48$

Alcohols:

$$\Delta H_a = 280.00 A + 12.61 \Delta^2\chi + 2.49 \chi_{PC}^4 - 73.33 \quad [32]$$

$r = 0.99999, s = 0.87, \varepsilon = 0.045\%, F = 73433226, n = 21$

where  $\varepsilon$  is the average relative error and  $F$  is the variance ratio. The regression statistics here are excellent. The more important aspect is the ability to make comparisons about the impact of structure differences on the heat of atomization. For example, the coefficient of  $A$ , which also indicates the added heat of atomization per methylene group, is essentially the same in alkanes and alcohols but somewhat less in alkenes. Effects of unsaturation are greater in alkylbenzenes than in alkenes, as reflected in the coefficient of  $\Delta^2\chi$ .

## Ionization Potential

Because the delta chi indexes,  $\Delta'''\chi$ , emphasize the role of the nonsigma electrons, they have become important in relating to properties which are more dependent on lone pair and  $\pi$  electrons. The delta chi indexes introduced above have also been found useful in an analysis of the ionization potentials of a set of 24 alkyl amines, alcohols, and ethers. The delta chi for the first three orders are defined as above.

For the ionization potential (IP) data obtained by Watanabe,<sup>120</sup> the following equations have been obtained<sup>3</sup>:

$$\text{IP(eV)} = 5.513 \Delta^0\chi + 6.595 \quad [33]$$

$$r = 0.808, s = 0.578, F = 41, n = 24$$

$$\text{IP(eV)} = 5.014 \Delta^0\chi + 5.166 \Delta^1\chi + 5.341 \quad [34]$$

$$r = 0.955, s = 0.299, F = 109, n = 24$$

$$\text{IP(eV)} = 5.364 \Delta^0\chi + 6.341 \Delta^1\chi + 1.517 \Delta^2\chi + 4.243 \quad [35]$$

$$r = 0.993, s = 0.123, F = 461, n = 24$$

The variable  $\Delta^0\chi$  in QSAR Eq. [33] provides discrimination only among the three molecular classes because the atom contribution to  $\Delta^0\chi$  is zero for saturated carbon atoms. The  $\Delta^0\chi$  provides very little structure information with respect to skeletal variation, but it does encode the atom identities and some of the skeletal environment immediately surrounding the heteroatom. The addition of the  $\Delta^1\chi$  variable greatly increases the discrimination among the three classes because it encodes skeletal information about the carbon atoms  $\alpha$  to the heteroatom. The QSAR is improved considerably by the addition of  $\Delta^1\chi$ . Finally, the addition of  $\Delta^2\chi$  further improves the QSAR by adding information about the broader reaches of the skeletal environment of each heteroatom, namely, atoms  $\beta$  to the heteroatoms. It can also be seen that the effects of atoms  $\beta$  to the heteroatom are much less important than the heteroatom itself or the  $\alpha$  carbon atoms. By the introduction of the delta-chi indexes, an atom level interpretation is made possible.

## Molar Refraction

A property for which generally good experimental data are available is molar refraction (MR), a composite of refractive index, density, and molecular weight. Because of its relation to polarizability, it has been used in many QSAR relations as a regression variable. It is observed that the relation between MR and skeletal variation is very different from the relation between heat of atomization and structure. The heat of atomization for a set of alkane isomers is generally ranked rather well by the  $^1\chi$  index; the most stable isomer in a set generally has the smallest value of the chi-one index. Such is not the case with MR; the MR values in an isomer set follow a much different pattern with  $^1\chi$ .<sup>2,3</sup> We expect, then, a rather different QSAR for MR than for  $\Delta H_a$ . For alkanes<sup>3</sup> the following QSAR is obtained:

$$\text{MR} = 3.832 {}^0\chi + 4.438 {}^1\chi - 0.8728 {}^3\chi_p - 0.4828 {}^4\chi_p - 0.4558 \quad [36]$$

$$r = 0.99999, s = 0.043, F = 194694, n = 55$$

This high quality correlation reveals the structure basis of the chi indexes and their ability to relate structure to this property.

It should be noted that chi path indexes depend on both number of atoms and skeletal variation. For example,  ${}^1\chi$  increases with number of atoms but decreases with skeletal branching, whereas  ${}^2\chi$  increases with both atom count and skeletal branching. In general,  ${}^m\chi_p = f(\text{atom count, skeletal branching})$ . It would be advantageous to separate these two molecular variations. For alkanes, this can be accomplished by defining a difference index, which is referenced to the chi index for the corresponding normal or unbranched index:

$$d{}^m\chi_p = {}^m\chi_N - {}^m\chi_P \quad [37]$$

where  ${}^m\chi_N$  is the chi index of order  $m$  for the unbranched alkane. In this manner,  $d{}^m\chi_p$  is found to be less dependent on atom count. Using indexes defined in this manner, it is possible to obtain higher quality QSARs for the alkane MR data.

$$\text{MR} = 4.643 A + 3.122 d^0\chi + 2.596 d^1\chi - 0.1193 {}^4\chi_{PC} + 2.093 \quad [38]$$

$$r = 0.99999, s = 0.034, F = 334134, n = 55$$

Using these chi difference indexes, it is possible to see the structure effects due to skeletal variation at each atom ( $d^0\chi$ ), skeletal effects due to branching but limited to the accumulated nearest neighbor effect ( $d^1\chi$ ), and due to branching in the skeleton ( ${}^4\chi_{PC}$ ). Further, because atom count and skeletal variation are uncoupled, it is possible to gain higher quality regression equations.

Chi indexes have also been used to obtain QSAR for mixed classes of compounds. Recently, Hall and Aaserud have reported an QSAR for a set of 46 alkylmonosilanes, 46 alkanes, and 51 polychloroalkanes.<sup>121</sup>

$$\text{MR} = 2.921 {}^0\chi^v + 4.837 {}^1\chi + 0.418 {}^2\chi^v + 0.781 \quad [39]$$

$$r = 0.9996, s = 0.369, F = 68300, n = 143$$

This is the first reported high quality QSAR for MR which includes hydrogen, carbon, silicon, and chlorine.

### QSAR of General Anesthetics

The anesthetic potencies of a set of hydrocarbons, ethers, and ketones were examined by DiPaolo.<sup>27</sup> The potency is represented as  $pC_{\text{eff}} = -\log(c_{\text{eff}})$ , where  $c_{\text{eff}}$  is the effective anesthetic concentration. The experimental error is reported as 0.17. In the analysis, DiPaolo reported that the dependence on molecular size was best given by  ${}^1\chi$  but that the relation is inversely proportional. Further, the valence chi path index of order four is required to provide the QSAR equation of high quality:

$$pC_{\text{eff}} = 8.539 / {}^1\chi + 1.487 {}^4\chi_p{}^v - 2.895 \quad [40]$$

$r = 0.943, s = 0.17, F = 97, n = 27$

The equation accounts for 89% of the variance in the potency, and the standard error is at the level of the reported experimental error. DiPaolo showed that a plot of potency versus  ${}^1\chi$  reveals nonlinear dependence with a maximum point. However, a quadratic form is not able to unite the data for all three classes of compounds into a single equation. It is the combination of the reciprocal of  ${}^1\chi$  together with  ${}^4\chi_p{}^v$  that gives a single equation for the three chemical classes. Examination of the equation indicates that the maximum occurs at  ${}^1\chi = 4.5$  and  ${}^4\chi_p{}^v = 0.863$ , values near to those for butyl propionate and propyl butyrate.

### Phenol Toxicity to Fathead Minnows

Molecular connectivity indexes have been widely used to develop QSARs for toxicity and other properties of environmental interest.<sup>9</sup> Hall and Kier investigated the toxicity of substituted phenols to fathead minnow (*Pimephales promelas*)<sup>48</sup> and have shown that a two-variable equation gives good account of the toxicity data.

$$pLC_{50} = 0.906 {}^3\chi_p{}^v + 0.205 {}^1\chi + 2.52 \quad [41]$$

$r = 0.934, s = 0.30, F = 75, n = 25$

where  $pLC_{50}$  is the negative log of the molar concentration to produce 50% kill of the fish in the sample. The fathead minnows were 30 day old fish subjected to varying concentrations of phenol compounds for 96 hr in flowthrough conditions.<sup>48</sup>

A question that sometimes arises in QSAR analyses deals with the possibility of random correlations. In this study the regression equations were carefully and extensively analyzed for spurious effects due to random correlations. The regressions were repeated with randomly selected observations deleted, and no significant effects were observed in the regression equations. Further, the process of equation selection was repeated using random numbers in place of the chi indexes. No correlation obtained with the random numbers was found to be significant in comparison to that of Eq. [41]. These random number analyses have also been carried out for other QSAR investigations.<sup>122,123</sup>

Once a QSAR relation has been developed, a structural interpretation can be given. The direct relation of  $pLC_{50}$  to  ${}^1\chi$  indicates that toxicity generally increases with an increase in molecular size for these data. It is, of course, not expected that such an increase goes to infinity, but in the data set there is no evidence of a turn-down in toxicity with increasing molecular size. The relation to the  ${}^3\chi_p{}^v$  index indicates two general features: (1) higher row heteroatoms

increase toxicity more than methyl groups, and (2) adjacent substitution patterns increase toxicity over nonadjacent patterns. A general result of Eq. [41] is that size, atom identity, and arrangement, in specific patterns, are responsible for toxicity. Number of atoms, per se, is insufficient for a good correlation. This effect can be well demonstrated by examining the correlation using only compounds containing chlorine in varying numbers and positions.

$$\text{pLC}_{50} = 0.950 \cdot {}^3\chi_p^\nu + 2.70 \quad r = 0.950, s = 0.23, F = 51, n = 8 \quad [42]$$

The corresponding correlation with number of atoms has  $r = 0.63$ .

### Inhibition of Microsomal *p*-Hydroxylation of Anilines by Alcohols

Sabljić has investigated a series of aliphatic alcohols which inhibit the *p*-hydroxylation of anilines by cytochrome P-450.<sup>124</sup> For the 20 alcohols studied, it was found that an inverse relation exists between the activity and the  ${}^0\chi^\nu$ . Addition of the  ${}^4\chi_{PC}$  index yields a satisfactory QSAR equation:

$$\begin{aligned} \text{pIC}_{50} &= -6.88 / {}^0\chi^\nu - 1.14 \cdot {}^4\chi_{PC} + 1.85 \\ r &= 0.983, s = 0.16, F = 249, n = 20 \end{aligned} \quad [43]$$

The activity is expressed as  $\text{pIC}_{50}$ , the negative logarithm of the concentration to cause 50% inhibition. The equation is statistically significant at the 99.5% level. Inhibition increases with increase in molecule size but decreases with increase in molecular branching as given by the  ${}^4\chi_{PC}$  index. Sabljić concluded that the inhibitory power is a fine balance between the size and the degree of branching of the alkyl chain. Based on his analysis, Sabljić made some mechanistic suggestions.

### Antiviral Activity of Benzimidazoles against Flu Virus

Several studies using molecular connectivity have been performed on antiviral data.<sup>3,4,9</sup> Tamm et al.<sup>125</sup> developed data for the antiviral activity of alkyl-substituted benzimidazoles against the Lee strain of the flu virus. Hall and Kier have analyzed that data using chi indexes.<sup>42</sup> It is found that the activity depends heavily on arrangement of substituents but not on the atom count. This analysis is revealed in the following QSAR.

$$\text{pI}_{75} = 1.40 \cdot {}^6\chi_p + 1.11 \quad r = 0.950, s = 0.17, F = 120, n = 15 \quad [44]$$

It was also shown that the  ${}^4\chi_p^v$  index could be used as a significant second variable in the QSAR equation. Hansch-type analysis yielded  $r = 0.90$ .<sup>126</sup>

Hall and Kier considered the possibility that more specific structure-activity information could be obtained by considering the subgraphs which compose the  ${}^6\chi_p$  index. Each chi index is a sum of contributing subgraphs from the whole molecule but not necessarily uniformly across the molecule. By considering path-six subgraphs from various identifiable parts of the molecule, Hall and Kier developed the following conclusions about the structure-activity relations in the data set: (1) substitution on the imidazole 2-position is more important to activity than other positions; (2) branched substituents on the 2-position should be higher in activity, especially those with a branch point  $\alpha$  to the ring; and (3) positions on the six-membered rings are not differentiated with respect to activity. This type of analysis of the structure contributions to activity is possible because the chi indexes are representations of molecular structure.

### Bioconcentration Factor for Phenyl and Biphenyl Compounds

Sabljić and Trinajstić<sup>9</sup> and Koch<sup>45</sup> reported investigations on data of environmental interest; Rouvray<sup>6</sup> reported a review. One property of environmental and biological interest is the bioaccumulation of organic chemicals in aquatic organisms. Sabljić has investigated several such data sets including a set of halocarbons (chlorinated hydrocarbons, benzenes, biphenyls, and diphenyloxides).<sup>40</sup> Sabljić demonstrated that the bioconcentration factor (BCF) for these compounds has a nonlinear relation with structure variables such as the number of atoms. It was found that the second-order valence chi index in a parabolic form gave good correlation with the data.

Hall and Kier reexamined this BCF data set using the response surface optimization<sup>127</sup> technique as reported for a neurotoxicity data set.<sup>128</sup> In this approach the nonlinear parabolic form is extended to a general two-variable parabolic form. The analysis can be performed using ordinary multiple linear regression programs or an extended form of the analysis can be performed using SAS.<sup>129</sup> For the 20 compounds investigated by Sabljić, Hall and Stewart<sup>130</sup> used the sum and difference of the zero order chi indexes,  ${}^0\chi$  and  ${}^0\chi^v$ , defined as follows:

$${}^0S = {}^0\chi + {}^0\chi^v \quad [45]$$

$$\Delta {}^0\chi = {}^0\chi - {}^0\chi^v \quad [46]$$

As indicated earlier in the discussion of  $\Delta^2\chi$  for the ionization potential study, the difference chi indexes encode the nonsigma electron information. The sum index has been shown to relate to molecular size.<sup>127</sup> Further, these sum and difference indexes are orthogonal. Using the sum/difference variables, the following results were obtained:

$$\log(\text{BCF}) = 0.795 {}^0S - 0.0170 ({}^0S)^2 - 0.530 \Delta {}^0\chi - 0.787 (\Delta {}^0\chi)^2 + 0.0632 ({}^0S \Delta {}^0\chi) - 4.735 \quad r=0.984, s=0.22, F=88, n=20 \quad [47]$$

This parabolic surface has a  $\log(\text{BCF})_{\max}$  at the value of 4.86 for the index values,  ${}^0S_{\max} = 24.568$  and  $\Delta {}^0\chi_{\max} = 0.650$ , which correspond to values for the zero order chi indexes:  ${}^0\chi_{\max} = 12.609$  and  ${}^0\chi^v_{\max} = 11.959$ . This maximum point is near the values for the tetrachlorobiphenyls. In addition to the phenyl and biphenyl compounds in the data set, the compounds DDT, heptachlor, and diel-drin also fit this model well.

### Physical Significance of Molecular Connectivity Indexes

In the connectivity method, as illustrated here, as well as in related topological methods, the approach is fundamentally different from traditional biological QSAR methods based on an assumed mechanism and using physicochemical properties as regression variables. The line of reasoning in topological methods runs directly from structure to activity, including biological activity, and not indirectly through an intermediate physical property. In this framework, the chi indexes represent molecular structure. The connectivity method seeks to extract the structure information from a data set and relate it directly to the set of activities. Viewed from this perspective, it is seen that the physical significance of molecular connectivity is *representation of molecular structure*.

Consider a simplified illustration of the foregoing QSAR examples. Consider a list of normal alkanes together with their water solubility and boiling points. A plot shows that solubility (in logarithmic form) is linear with number of carbon atoms and that boiling point is nonlinear. Such a relation is a QSAR based on the simple structure feature, number of carbon atoms. A linear equation captures all the structure information available in this data set. (The structure information could, of course, be represented in other ways, such as number of methylene groups, number of hydrogen atoms, number of carbon-carbon bonds, etc.) It is important to note here that no assumption has been made about the relation between water solubility and number of carbon atoms. This is an example of what Adamson has called a mechanism-free model.<sup>131,132</sup>

In similar fashion, a QSAR can be developed for the boiling point, although the mathematical relation is more complicated, perhaps a logarithmic form. In both cases, however, one may proceed to use the mathematical equations to interpolate and extrapolate or to attempt to invert the process and determine the molecular structure that corresponds to a given property value. It is clearly seen here that the physical significance of the regression correlate, "number of carbon atoms," is molecular structure. That is, what is "encoded" in the number of carbon atoms adequately expresses what we know about the normal alkanes in terms of molecular structure for this particular data set.

Further, it is not at all strange that we can use "number of carbon atoms" to express the structure relation for many different properties. It is not being said that "number of carbon atoms" is in any way synonymous with solubility or boiling point or, in fact, that "number of carbon atoms" stands for solubility or boiling point. The structure-activity model essentially represents the relationship between structure and property in a quantitative mathematical form suitable for further use.

The molecular connectivity indexes represent molecular structure in a manner analogous to the count of carbon atoms, but in a much more general way. That is, chi indexes are weighted counts of structure features with the same mathematical qualities as counts, but with much more structure information.

---

## CHARACTERIZATION OF MOLECULAR SHAPE

The concept of the shape of a molecule is an important, although elusive, notion for the chemist. It has proven difficult to make quantitative, although there have been various attempts. The familiar model kits and graphic depictions indicate the widespread nature of the shape notion. Molecules are variously described as flat or spherical or asymmetric. Very specific terms may be applied, such as those derived from symmetry, based on spectroscopic or crystallographic methods. It is a common practice to represent molecules as objects with certain sizes and boundaries. Accordingly geometric analyses and computer graphic representations are popularly invoked to describe and to seek relationships between molecular shapes and their chemical and biological events. Some have questioned the validity of this model.<sup>133</sup>

From the quantum mechanical point of view, molecules are not hard objects with well-defined boundaries.<sup>134</sup> For certain applications, boundary surfaces are chosen, somewhat arbitrarily, to provide a basis for making certain calculations or visualizations. It is questionable whether the word shape is appropriate, if it is intended to imply a surface. The information derived from spectroscopic data most appropriately applies to the potential function of the molecule, where the term shape may be used but not in the same manner as applied to a hard object with well-defined surfaces. It might be more appropriate to use other terms, such as framework, which is defined as a structure serving to hold the parts of something together. The word steric configuration, having to do with the spatial arrangement of atoms in a molecule, has been used, but it is also incomplete.

Further complications in quantitating molecular shape arise from the variable nature of atom relationships across space. This conformational variability presents uncertainty in depicting a reliable shape, although quantum mechanics addresses the problem of energy-preferred structures. A molecule, as

an ensemble of atoms, is highly irregular in a geometric sense, further complicating meaningful description and quantitation.

Bearing these considerations in mind, nonetheless, shape quantitation presents a challenge well worth the effort. The potential value of shape quantitation is clearly demonstrable in physical and biological studies where this attribute is influential. Accordingly, we have addressed this problem and have contributed one possible approach to its solution.

---

## BACKGROUND: STERIC OR SHAPE INFLUENCE

In chemical phenomena, two classes can be described as being influenced by steric effects. First, groups of atoms affect the reactivity of a nearby part of a molecule. This steric effect frequently manifests itself as a repulsive influence toward a reagent attacking a nearby reaction center. These effects are analyzed and encoded into substituent indexes from a standard reaction, based on their relative inhibitory effect.

In the second class of phenomena are intermolecular interactions, where a certain degree of complementarity must be achieved for a reaction to occur. There must be a "fit" or "recognition" between molecules or between molecule and receptor, governed by the shape of each. The degree of fit or complementarity (similarity) has led to indexes encoding shape.

Studies of these two facets of steric influence have spawned different approaches to the quantitation of this structural feature. A brief review of some methods illustrates the progress and problems in this area.

---

## METHODS FOR STERIC QUANTIFICATION

### Quantitation of Influence on Properties

Quantitation of the effect of the shape change of part of a molecule is based on the change in the measured value of a physical or chemical property. In particular, Taft<sup>136</sup> and others recognized that the hydrolysis rates of esters are influenced by the steric (or bulk) interference of parts of the molecule near the reactive site. In a specific case, the acid hydrolysis rates of esters are viewed as being governed largely by the influence of the acyl moiety. Its size, bulk, or, in general, its steric effect provided an inhibitory influence upon the attacking reagent. Using ethyl acetate (where the substituent on the acyl moiety is  $-\text{CH}_3$ ) as a standard reference, the rates of hydrolysis of esters formed by replacing  $-\text{CH}_3$  with another

group, R, were measured. The logarithm of the relative rates is thus one measure of the steric effect of a group R. These substituent steric parameters, Taft  $E_\alpha$  values, have been widely used to reflect this structural aspect of a group. Over the years, several analyses and refinements have been made.<sup>136,137</sup>

The assumptions inherent in this approach are that the reaction is free from variation in electronic (inductive) effects through bonds and is occurring with the same mechanism in all cases studied. It is obvious that some groups cannot be directly evaluated from the ester hydrolysis reaction. Other reactions have been used, but the interrelation of parameters from different reactions is not always clear.

## Geometric Models

When a molecule is represented as an object with well-defined boundaries, geometric methods of analysis may be applied. Several steric parameters have been derived from the model in which atoms, hence molecules, are represented as geometric objects with surfaces and volumes. The "boundary" of an atom is usually determined by its van der Waals radius. Surface area and volume of groups and molecules can be estimated. Bondi,<sup>103</sup> Hermann,<sup>138</sup> and Pearlman<sup>139</sup> are among those who have studied surface area and volume. Molar refractivity (MR), calculated from density and refractive index, has frequently been used to estimate the bulk or global volume of a molecule or group. A more recent trend is to use MR as a model for dispersion interactions.

Verloop<sup>140</sup> and co-workers have used scale models of molecules and groups to define the dimensions of a box which would enclose their most probable shape. A convention is adopted beginning with the selection of the longest axis of the structure followed by the designation of lateral axes. These dimensions, in concert or taken separately, have been used as shape parameters in QSAR analyses.<sup>140</sup>

## Object Comparisons

In another approach, also based on geometric methods, the silhouette or the molecular spatial extent has been compared with that of some reference molecule, noting coincident and noncoincident parts of the models. Simon and Szabada<sup>141</sup> introduced such a method called minimal topological difference (MTD) and a refinement called minimal steric difference (MSD). A reference pattern of atoms is deduced from available structure-activity data. Comparisons are then made with candidate molecules in terms of overlapping and non-overlapping molecular features. A scoring is made and used as a parameter reflecting relative similarity. A similar approach has been introduced by Hopfinger, who also utilized a reference compound for comparison.<sup>142</sup> Jurs has also contributed to this method.<sup>143</sup>

## Structure Description Based on Topology or Chemical Graph Theory

An alternative to a purely geometric approach is to view the molecular structure as possessing information, some of which relates to shape. The belief that molecular information is encoded in a graph is an old one. It is common for the organic chemist to represent portions of large molecules with hydrogen-neglected "skeleton" structures. The idea that graphs could be analyzed to provide information-rich indexes is first attributed to Wiener.<sup>69,70</sup> Since that time, several indexes have been derived from molecular graphs. Specifically, contributions have been made by Altenburg,<sup>76,77</sup> Gordon and Scantlebury,<sup>78</sup> Hosoya,<sup>79</sup> Randić,<sup>10</sup> Bonchev and Trinajstic,<sup>144</sup> Balaban,<sup>1,145</sup> and Kier and Hall.<sup>3</sup>

The graph representation of molecular structure together with the extraction of chemical information may be called chemical graph theory. The common characteristic in all of these graph-based approaches is the counting of elements of the graph. These elements are the graph vertexes, representing atoms, and graph edges, representing bonds, along with various combinations of graph edges. The information contained in these counts has been shown to relate closely to relative property values of molecules. The graph-based methods are clearly attempts to describe the structure of a molecule.

Attempts at quantitating molecular structure using shape-effect and geometry-based methods depict the molecule as having a surface with only incidental interest in the interior. In contrast, graph-based methods depict the molecule as being primarily a framework (the presence of atoms and their adjacency relationships) with no assumptions about a surface. Each has a rich potential for encoding structural information, and both should be explored for that potential.

Graph-based structure analyses had not been used for the specific objective of encoding relative shape. There was the frequent inference that indexes derived from these methods encode shape information, although none had been conceived with this specific objective in mind. It is from the graph of a molecule that we have derived the kappa indexes with the objective of encoding relative shape.<sup>59-61</sup>

---

## MODEL OF MOLECULAR SHAPE BASED ON CHEMICAL GRAPH THEORY

### General Model

The basic assumption in this model is that the shape of a molecule is a function of the number of atoms and their adjacency relationships (bonding scheme). The pattern of atoms and bonds is represented by a molecular skeleton or graph in which all hydrogen atoms are excluded, called the hydrogen-

Table 1 Counts of Graph Paths of Length 1, 2, and 3 for Hexanes

	A	$^1P$	$^2P$	$^3P$
	6	5	4	3
	6	5	5	4
	6	5	5	3
	6	5	6	4
	6	5	7	3
	6	6	6	6
	6	7	11	13
	6	7	10	14

suppressed graph. In Table 1 several isomers of hexane are shown, including three cyclic hexanes.

For our present purposes, attention is focused on counts of four graph quantities: the number of atoms, A; the number of bonds, using the term  $^1P$  to denote paths of one bond length; the number of two-path (two adjacent bonds) fragments,  $^2P$ ; the number of three-path fragments,  $^3P$ . In general for contiguous paths of length  $m$ ,  $^mP$  is the symbol used.

It is generally true that the shape of each molecule in an isomeric series is different. The most primitive count, the count of atoms, gives a broad classification as hexanes, but provides no useful information within this set. Clearly, to capture information about differences in structures possessing different shapes, we must use the path counts as information sources to derive an index. It is also anticipated that a single index will not encode all shape information. A manifold of indexes must be derived that carries information about different attributes of molecular shape. A summation of attributes could ultimately lead us to useful descriptions of this molecular property.

The counts of each order of path length can be viewed as describing individual attributes of shape, each a part of the manifold of attributes into which shape may be dissected. The use of path counts has early origins in graph theory-based structure index development. Beginning with the pioneering work

of Wiener,<sup>69,70</sup> contributions were made by Altenburg,<sup>76,77</sup> Balaban,<sup>1,145</sup> and Randić,<sup>10</sup> among others.

In the approach leading to kappa shape indexes, it is necessary to transform  ${}^mP$  into an index that carries information for any number of atoms in the molecule. To accomplish this, we define a particular shape attribute as having an intermediate relationship between two extreme shapes, each of which may be defined both pictorially and numerically. We take the position, for the present, that these extreme shapes must be common to subsets of molecules of any number of atoms. The extremes selected for any order of attribute,  $m$ , are the maximum,  ${}^mP_{\max}$ , and minimum,  ${}^mP_{\min}$ , counts of paths in the molecular graphs of molecules with a common atom count. A shape attribute of a particular order,  $m$ , for a particular molecule,  $i$ , is therefore

$${}^mP_{\max} \geq {}^mP_i \geq {}^mP_{\min} \quad [48]$$

where the number of atoms is the same for all three structures. This set of numerical relationships will be transformed into a single number for each attribute. To accomplish this, we examine each order and derive an algorithm to encode the shape information.

### First-Order Shape Attribute

For this attribute of shape, described by the counts of one bond fragments,  ${}^1P$ , we selected  ${}^1P_{\max}$  to be the complete graph, where all atoms are bonded to each other. For any number of atoms,  $A$ , the value of  ${}^1P_{\max} = A(A - 1)/2$ . For the  ${}^1P_{\min}$  structure, we have selected the linear graph where the value is  ${}^1P_{\min} = A - 1$ . In Table 2, entry No. 1 is the graph of  ${}^1P_{\max}$ , where  $A = 6$ . Entry No. 2 is the graph of  ${}^1P_{\min}$  where  $A = 6$ .

The shape attribute of the first order lies between the complete graph and the linear graph. This is the basis of our definition of this shape attribute. We are not considering, or numerically defining, spheres, ellipsoids, or other geometric figures.

We proceed to derive an algorithm which yields a numerical index for any molecule with  $A$  atoms and with  ${}^1P_i$ . To embrace the information for the extremes in Eq. [48], the products of ratios of  ${}^1P_i$  to  ${}^1P_{\max}$  and  ${}^1P_i$  to  ${}^1P_{\min}$  are used. We write down an index of shape of order one, using  ${}^1\kappa$  (kappa) as the index symbol:

$${}^1\kappa = 2 {}^1P_{\max} {}^1P_{\min} / ({}^1P_i)^2 \quad [49]$$

The "2" in the numerator is a scaling factor that makes  ${}^1\kappa = A$  when there are no cycles in the graph of the molecule. The index,  ${}^1\kappa$ , can be expressed in terms of the count of atoms,  $A$ :

$${}^1\kappa = A(A - 1)^2 / ({}^1P_i)^2 \quad [50]$$

**Table 2** Graphs of Structures Showing Paths Counts of Orders 1, 2, and 3 Illustrating  ${}^mP_{\max}$  and  ${}^mP_{\min}$ 

	<i>A</i>	${}^1P$	${}^2P$	${}^3P$
	6	15		
	6	5		
	7		15	
	7		5	
	8			9
	8			5
	9			12
	9			6

### Second-Order Shape Attribute

The second-order shape attribute is defined by the count of two-bond paths,  ${}^2P_i$ , and is related to the shape extremes represented by  ${}^2P_{\max}$  and  ${}^2P_{\min}$ . For  ${}^2P_{\max}$  we adopt the star graph, in which all atoms but one are adjacent to a central atom. This graph for  $A = 7$  is shown in Table 2, No. 3. The numerical value of  ${}^2P_{\max}$  for any count of  $A$  is  ${}^2P_{\max} = (A - 1)(A - 2)/2$ . For the other second-order shape attribute extreme,  ${}^2P_{\min}$ , we use the linear graph, shown in Table 2, No. 4, where  $A = 7$ . In general, for any value of  $A$ ,  ${}^2P_{\min} = A - 2$ .

An algorithm expressing this second-order attribute can now be written down using the same formalism as for  ${}^1K$ , a product of ratios. A second-order shape index,  ${}^2K$ , is defined:

$${}^2K = 2 \cdot {}^2P_{\max} \cdot {}^2P_{\min} / ({}^2P_i)^2 \quad [51]$$

The scaling factor of 2 in the numerator makes the value  ${}^2K = A - 1$  for all linear graphs, where  $A - 1$  is the number of graph edges of skeletal bonds for acyclic molecules. Equation [51] can be expressed in terms of the count of atoms,  $A$ :

$${}^2K = (A - 1)(A - 2)^2 / ({}^2P_i)^2 \quad [52]$$

### Third-Order Shape Attribute

The count of paths of three contiguous bonds,  ${}^3P$ , forms the basis for the description and quantitation of another shape attribute. This structural attribute is compared to two extreme structures defined by  ${}^3P_{\max}$  and  ${}^3P_{\min}$ . For the third-order attribute,  ${}^3P_{\max}$  is taken from a twin star structure shown in Table 2, No. 5 for  $A = 8$ , and No. 7 for  $A = 9$ . For  ${}^3P_{\min}$ , the linear graphs, Table 2, No. 6 and No. 8 are the corresponding representations. In general, for any odd value of  $A$ ,  ${}^3P_{\max} = (A - 1)(A - 3)/4$  and for any even value of  $A$ ,  ${}^3P_{\max} = (A - 2)^2/4$ . In general,  ${}^3P_{\min} = A - 3$ . A suitable algorithm in which third-order shape information,  ${}^3K$ , can be calculated for  ${}^3P_i$  is

$${}^3K = 4 {}^3P_{\max} {}^3P_{\min} / ({}^3P_i)^2 \quad [53]$$

The scaling factor of 4 is used in the numerator to bring the  ${}^3K$  onto approximately the same numerical scale as the other kappa values.

The  ${}^3K$  values can be expressed in terms of  $A$  using two equations:

$${}^3K = (A - 1)(A - 3)^2 / ({}^3P_i)^2 \quad \text{when } A \text{ is odd} \quad [54]$$

$${}^3K = (A - 2)(A - 3)^2 / ({}^3P_i)^2 \quad \text{when } A \text{ is even} \quad [55]$$

### A Shape Index from Zero-Order Paths

Based on path counts, it may be also possible to define a zeroth order index,  ${}^0K$ . It follows from our path count development that  ${}^0K$  should be derived from  ${}^0P$  fragments, a quantity that is, in fact, the count of atoms in a molecular graph. One attribute of an atom that should influence the shape of a molecule is the topological uniqueness of that atom within the molecule. One collective effect of atom topological uniqueness is the symmetry or redundancy of the molecule.

Atom uniqueness or redundancy is certainly a shape attribute that must play some role in the influence of structure on function. One approach to the quantitation of uniqueness is the use of the Shannon equation for information content,<sup>146</sup> which has been studied quite thoroughly by Brillouin<sup>147</sup> and Bonchev.<sup>148</sup> Kier has made use of the equation to encode molecular uniqueness, or "negentropy," and to relate to biological and physical properties.<sup>149</sup>

Two interrelated values may be derived from the Shannon equation. The first of these is  $I$ , the information content per atom:

$$I = - \sum p_i \log (p_i) \quad [56]$$

where  $p_i$  is the probability of randomly selecting an atom of a given type from the whole. The information content in the entire molecule with  $A$  atoms is  $IA$ .

Table 3 Selected Graphs to Illustrate the Range of Kappa Index Values

Kappa	A	Graphs/Kappa Value				
$^1\kappa$	6	0.667	2.344	3.061	4.167	6.000
$^2\kappa$	7	0.667	1.240	2.344	4.167	6.000
$^3\kappa$	9	2.000	2.880	4.500	5.878	8.000
$^0\kappa$	6	0.000	1.659	2.863	4.067	

Brillouin<sup>147</sup> has modified this expression to give the redundancy within the molecule,  $R$ :

$$R = 1 - (I/\log A) \quad [57]$$

We propose a new redundancy index,  $\Sigma(m_i)^{-0.5}$ , where  $m_i$  are the multiplicities within each topologically equivalent subset. A shape index,  $^0\kappa$ , could be any of these uniqueness and redundancy indexes in an effort to encode information about shape.

## SHAPE INFORMATION IN THE KAPPA VALUES

By using the limit-structure model for each order of shape attribute as we have done above, we can quantitate these attributes from their relationships to two extreme structures associated with each model.

In Table 3, a range of  $^1\kappa$  values is shown for structures ranging from a linear structure to one with a maximum number of cycles. The structural information encoded in  $^1\kappa$  is related to the complexity, or more precisely, the "cyclicity" of a molecule. All acyclic structures with six atoms would have  $^1\kappa = 6.000$ , or, in general,  $^1\kappa = A$  for acyclic structures.

In the case of  $^2\kappa$ , the structures and index values in Table 3 for  $A = 7$  reveal that information is encoded relating to the degree of star graph-likeness and

linear graph-likeness. Stated in more general terms,  ${}^2\kappa$  encodes information about the spatial density of atoms in a molecule.

The information in  ${}^3\kappa$  indexes can be illustrated by the set of structures in Table 3. The  ${}^3\kappa$  values are larger when branching is nonexistent or when it is located at the extremities of a graph. We can say that  ${}^3\kappa$  encodes information about the centrality of branching.

Finally, in Table 3 the influence of several structures on the value of  ${}^0\kappa$ , which we compute as  ${}^0\kappa = IA$  from the carbon skeleton is shown. The greater the topological symmetry, the lower the  ${}^0\kappa$  value. Stated another way, the greater the uniqueness of atoms, information content, or negentropy, the larger the value of  ${}^0\kappa$ .

---

## ENCODING ATOM IDENTITY

Up to this point, molecular graphs have been written down as though each atom is identical. Shape attributes have been interpreted on this basis. For application to chemical and biological problems, account must be made of atom differences.

Consider the pair of molecules, heptane and dipropyl ether. By any criteria, we would say that the shapes were very similar, though not identical. As derived so far, each  ${}^1\kappa$ ,  ${}^2\kappa$ , and  ${}^3\kappa$  value would be the same for this pair of molecules, implying shape identity. An additional consideration is necessary if we are to quantitate shape attributes where there are major differences in atoms and/or hybrid states. An approximation of shape equality between cyclohexane and benzene is probably not adequate for many structure-activity analyses. Some account must be taken of the fact that different atoms make different size contributions to a molecule, thereby influencing its overall shape.

### Modified Atom Count

Encoding atom identity can be accomplished in several ways. We have elected to modify the atom count,  $A$ , in Eqs. [50], [52], [54], and [55]. The modification is based on the awareness that a non-C( $sp^3$ ) atom, counted in arriving at the value of  $A$  for the molecular graph, is contributing more or less than a C( $sp^3$ ) contribution to the shape. Therefore, that particular atom should be counted more or less than 1, the increment or decrement called  $\alpha$ , which is based on the size contribution of the atom in question relative to C( $sp^3$ ).

One basis for comparing the size contribution of atoms is to use some function of the ratio of atomic radii. The question is, should we compare radii, surface area, or volumes of atoms? This would be reflected in a ratio of radii to

**Table 4** Alpha Values from Covalent Radii<sup>60</sup>

Atom Valence State	$r$ (Å)	$\alpha$
C (sp <sup>3</sup> )	0.77	0.0
C (sp <sup>2</sup> )	0.67	-0.13
C (sp)	0.60	-0.22
N (sp <sup>3</sup> )	0.74	-0.04
N (sp <sup>2</sup> )	0.62	-0.20
N (sp)	0.55	-0.29
O (sp <sup>3</sup> )	0.74	-0.04
O (sp <sup>2</sup> )	0.62	-0.20
F	0.72	-0.07
P (sp <sup>3</sup> )	1.10	0.43
P (sp <sup>2</sup> )	1.00	0.30
S (sp <sup>3</sup> )	1.04	0.35
S (sp <sup>2</sup> )	0.94	0.22
Cl	0.99	0.29
Br	1.14	0.48
I	1.33	0.73

the first, second, or third power. In initial studies on this problem we have adopted a ratio of covalent radii<sup>60</sup>:

$$\alpha_x = (r_x/r_{C(sp^3)}) - 1 \quad [58]$$

where  $\alpha$  represents a single parameter or a sum of parameters for several non-C(sp<sup>3</sup>) atoms.

The denominators of the general Eqs. [50], [52], [54], and [55] can be expressed in terms of  $A$ . Because  ${}^{m\prime}P_{max}$  and  ${}^{m\prime}P_{min}$  can be expressed in terms of  $A$ , then any value of  ${}^{m\prime}P_i$  can also be expressed as a function of  $A$ . If a molecule with non-C(sp<sup>3</sup>) atoms is described by  $(A + \alpha)$ , where  $\alpha$  is the sum of the  $\alpha_x$ , for all the atoms in the molecule, then the denominator is taken as  $({}^{m\prime}P_i + \alpha)^2$ . Rewriting the general expressions for the kappa indexes modified by  $\alpha$ , we have for  ${}^{m\prime}\kappa_\alpha$ :

$${}^1\kappa_\alpha = (A + \alpha)(A + \alpha - 1)^2 / ({}^1P_i + \alpha)^2 \quad [59]$$

$${}^2\kappa_\alpha = (A + \alpha - 1)(A + \alpha - 2)^2 / ({}^2P_i + \alpha)^2 \quad [60]$$

$${}^3\kappa_\alpha = (A + \alpha - 1)(A + \alpha - 3)^2 / ({}^3P_i + \alpha)^2 \quad A \text{ is odd} \quad [61]$$

$${}^3\kappa_\alpha = (A + \alpha - 2)(A + \alpha - 3)^2 / ({}^3P_i + \alpha)^2 \quad A \text{ is even} \quad [62]$$

A summary of  $\alpha$  values is found in Table 4.

Table 5  ${}^2\kappa_\alpha$  Values for Selected Molecules

	5.000		0.991
	6.000		1.606
	1.440		1.553
	1.633		1.500
	2.222		3.930
	3.061		4.290
	4.000		4.480
	4.740		

### Effect of Alpha Inclusion in Kappas

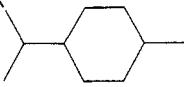
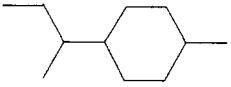
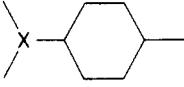
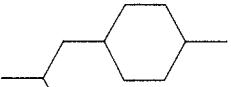
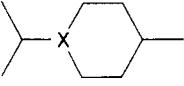
To illustrate briefly the information in the alpha-modified kappa values,  ${}^2\kappa$  is used as an example. Table 5 shows the  ${}^2\kappa_\alpha$  values for several molecules. The butyl halides in the list reveal the influence of the covalent radius on the shape modeled by the  $\alpha$  parameter on the  ${}^2\kappa_\alpha$  value. As expected, in bromobutane  ${}^2\kappa_\alpha$  is the largest, paralleling a shape that is less star graph-like than the other two halides. The same trends are noted in the cyclic molecules in Table 5.

The algorithms for  ${}^1\kappa_\alpha$  and  ${}^2\kappa_\alpha$  do not distinguish which atom or atoms are to be modified by  $A + \alpha$  within the molecule. The insertion of an  $A + \alpha$  designated atom anywhere within a molecule leads to the same calculated  ${}^1\kappa_\alpha$  or  ${}^2\kappa_\alpha$  values. The inference is that the shape attribute encoded by  ${}^1\kappa_\alpha$  or  ${}^2\kappa_\alpha$  indexes is the same regardless of the position(s) of the non- $C(sp^3)$  atom(s).<sup>61</sup>

To illustrate this point, in Table 6 the graph is shown that represents the all  $C(sp^3)$  molecule 1-methyl-4-isopropyl cyclohexane. For this molecule, the value of  $\alpha = 0$  is used when calculating  ${}^2\kappa$ . The structure is now modified by inserting at various positions a non- $C(sp^3)$  atom, X, which has a value of  $\alpha = 1.0$ . The choice of  $\alpha$  here is the equivalent of inserting an additional  $C(sp^3)$  atom into the original structure.

The values of  ${}^1\kappa_\alpha$  and  ${}^2\kappa_\alpha$  are the same for each isomer depicted in columns 1 and 2 of Table 6. The result is that the shapes of all molecules described by the same  ${}^1\kappa_\alpha$  or the same  ${}^2\kappa_\alpha$  are the same, regardless of where the actual modified atom is. This is not true in the case of the calculated  ${}^3\kappa_\alpha$  values. For this index, the  $\alpha$  modification of A used in calculating a  ${}^3\kappa_\alpha$  value may lead

**Table 6** Structural Isomers Illustrating Effect of Heteroatoms on  ${}^m\kappa_a$  Values

Heteroatom Position	$\alpha_x = 1.0$			
	All-C(sp <sup>3</sup> ) Isomer	${}^1\kappa_a$	${}^2\kappa_a$	${}^3\kappa_a$
		9.091	4.133	2.500
		9.091	4.133	3.265
		9.091	4.133	2.844

to different values of the index, depending upon the position of atom X in the molecule. The  ${}^3\kappa_a$  index must be calculated and used with this result in mind.

### KAPPA INDEX VALUES FOR SMALL MOLECULES

In some small molecules, certain of the  ${}^mP$  quantities may not be defined or are considered to be zero. This presents problems in applying the kappa algorithm. The calculation of a  ${}^1\kappa$  value is possible for any molecule except those represented by a single point, i.e., methane. In general, for a straight chain molecule,  ${}^1\kappa = A$ , and so an extrapolated value of  ${}^1\kappa = 1.000$  is adopted for methane.

The calculation of  ${}^2\kappa$  values leads to nonzero values in all cases except for any graph representation of one or two atoms, such as methane and ethane. In both cases,  ${}^2\kappa$  from Eq. [52] is zero. In general, for straight chain molecules,  ${}^2\kappa = A - 1$ . Values of  ${}^2\kappa = 1.000$  for ethane and  ${}^2\kappa = 0$  for methane are extrapolated and proposed as useful in cases where these molecules are part of a structure-activity analysis.

In the case of  ${}^3\kappa$  values calculated from Eqs. [54] and [55], the values for methane, ethane, and propane are zero, whereas the value for butane is 4.000,

the same as for pentane. By linear extrapolation, more useful  ${}^3\kappa$  values for these molecules are derived: methane  ${}^3\kappa = 0$ , ethane  ${}^3\kappa = 1.450$ , propane  ${}^3\kappa = 2.000$ , and butane  ${}^3\kappa = 3.378$ . These estimates are used to provide numerical values for small molecules that may appear in series being analyzed.

---

## MOLECULAR SHAPE QUANTITATION

### General Model

Within this paradigm, we view the shape or steric configuration of a molecule as being a function of a collection of attributes, each attribute being a function of the path count of a particular order. If the function is a summation, then molecular shape may be represented as

$$\text{Molecular Shape} = a {}^0\kappa + b {}^1\kappa + c {}^2\kappa + d {}^3\kappa + \dots \quad [63]$$

The summation is left open because we believe that many attributes define a molecular shape.

### Higher Order Indexes

After reviewing this concept based on a kappa series representing shape, a question arises: what about a possible  ${}^4\kappa$  or higher order index? Our preliminary studies reveal that the limiting structures for  ${}^4\kappa$  and higher indexes do not maintain their structural integrity, that is, the limiting structure for  ${}^4P_{\max}$  changes as the atom count increases. The interpretation based upon a maximum structure with common attributes independent of atom count is not sustained; hence the meaning of  ${}^4\kappa$  and higher indexes is not atom-count invariant. At present we have not extended the kappa index concept beyond order 3.

### Additivity

A property observed with the  ${}^2\kappa$  index is that values for fragments can be summed to approximate closely the calculated  ${}^2\kappa$  value for the combined structure. The implication is that the index is additive in regard to the information generated. The quality of the estimation of the  ${}^2\kappa$  value for a molecule from a sum of fragment  ${}^2\kappa$  values is a function of the fragments used. Put another way, we can dissect a molecule into pairs of fragments of varying quality based on their ability to reflect the shape of the original molecule. Use of this property is explained later in a discussion of group shape values.

---

## GENERAL APPLICATIONS

### Shape Similarity

One application of kappa indexes is quantitation of shape similarity in contrast to the absolute quantitation of size. The kappa values permit a rational prediction of which molecules have a high degree of shape similarity. Thus, *n*-pentane, diethylether, and diethylamine have similar kappa values. Electronic structure is information to be encoded into other indexes. The kappa values predict that these three molecules have very similar shape.

A comparison of cyclic and acyclic  ${}^2\kappa$  values is illuminating. The values for cyclohexane and 2,3-dimethylbutane are identical. Two-dimensional traces of Dreiding models of these two molecules reveal a high degree of superimposability. This is a consequence of comparable shape, from  ${}^2\kappa$  values, and comparable size, from  $A$  values. Identical  ${}^2\kappa$  values for durene (1,2,4,5-tetramethylbenzene) and naphthalene correspond to a high degree of superimposability, due a similar shape and size ( $A = 10$  for both).

With all orders of kappa values, a more comprehensive description of molecular shape is possible. This should permit a useful quantitation of shape similarity.

### Cavity Definition

A second application of the  ${}^2\kappa$  index is its use to predict candidate molecules to fill molecular cavities. With the increasing use of molecular graphics, the fit, docking, or intercalation of molecules into cavities in macromolecular simulations becomes an important consideration in drug design. The visualizations of proposed receptor sites, enzyme active sites, and other cavities and spaces of interest in macromolecules make it possible to make measurements of the dimensions of a cavity. Of course, the validity of these images depends on the quality of the input data and the assumptions attending the calculations. If the visualized details of a cavity are to be believed, then there is certainly some interest in what molecules may fit that cavity or some part of it.

The dimensions of a computer-visualized cavity are readily accessible. From these dimensions a generalized shape may be quantitated and a prediction made as to which molecules should be considered as candidates for fitting the cavity. To transfer information from linear measurements of cavity dimensions to  ${}^2\kappa$  values of shape, some approximate relationship must be established between these two metrics. The most convenient approach is to recognize that a cavity is of irregular shape; hence rectangular dimensions are not universally appropriate. We can, however, take advantage of the fact that  ${}^2\kappa$  values for cycles are very similar to  ${}^2\kappa$  values of acyclic molecules that have similar shapes on a two-dimensional trace. The identical  ${}^2\kappa$  values for cyclohexane and 2,3-

dimethylbutane demonstrate this generalization. This shape congruence can be used to define a cavity in terms of the length of its perimeter, expressed as the number of carbon atoms in a cyclic alkane which can be laid out in contact with the walls of the cavity. This is a two-dimensional estimate of the cavity perimeter, and so a particular plane must be selected to be used in an analysis of cavity shape. The perimeter would likely be defined in terms of the van der Waals contact distance separating the atoms in the cavity walls from the atoms in the cyclic alkane being used to measure the cavity perimeter.

From the dimensions of a cavity, an estimate can be made of the number of carbons,  $C_i$ , in a cyclic alkane that can be accommodated by the cavity with a perimeter of  $X$ , expressed in Å. A good approximation is

$$C_i = X/3 \quad [64]$$

If more than one plane of the cavity is to be considered, additional  $C_i$  estimates would be made.

From each  $C_i$  value, we calculate a  ${}^2\kappa$  index, describing the shape of the cavity in one plane. In general, the  ${}^2\kappa$  value of a cycle is

$$(A - 1)(A - 2)^2/A^2 \quad [65]$$

With the shape of one plane of the cavity expressed in terms of  ${}^2\kappa$ , it is now possible to predict which molecules would be good candidates to fit within the cavity. For example, Shen<sup>150</sup> studied analogs of indomethacin and has hypothesized a receptor accommodating the principal features of this molecule to bring about antiinflammatory activity. One part of that receptor was a flat area 6 by 5 Å, which interacted with the indole ring as substituted in the indomethacin molecule. This flat receptor area of Shen has an estimated perimeter of 22 Å. This would accommodate a cycle of 7.333 carbons from Eq. [64]. The shape index of this flat area is calculated from Eq. [65] to be  ${}^2\kappa = 3.35$ . The fragment of indomethacin fitting this receptor feature has a calculated  ${}^2\kappa$  value of 3.293, in close agreement with the receptor feature shape index.

Among active antiinflammatory agents other comparisons are equally satisfying. The corresponding fragment of methylcyclohexyl-2-phenylpropionic acid has a calculated  ${}^2\kappa$  of 3.240, whereas the corresponding value for the fragment of phenylnaphthalene acetic acids is 3.395. Similar compounds that are not active antiinflammatory agents have a fragment with a  ${}^2\kappa$  value of 3.539, higher than the postulated receptor area. The definition of cavities in this way may be used to screen candidate molecules for approximate fit at active sites.

## Molecular Flexibility

The terms flexibility and its antonym rigidity appear in discussions of structural attributes influencing chemical and biological events. Related terms

include rotational freedom, segmental motion, and mobility. A general definition of flexibility, "capable of conforming to changing or new situations; yielding to influence," is not useful to permit the quantitation of this chemical phenomenon. Indeed a quantitation has been an elusive quest since the concept of conformational variability was added to our understanding of molecular structure.

Luisi addressed this problem and recognized that the phenomenon can be dissected into two concepts.<sup>152</sup> In the first case, flexibility describes an equilibrium in which the molecule exists in a small number of conformations relative to all those conceivably based on the molecular structure. This thermodynamic conformational flexibility describes a situation in which many rapidly interconverting conformations exist. The second concept relates to the rate of interconversion among different conformations; hence, this is a kinetic conformational flexibility related to the energy barriers between conformations. The potential energy barriers govern the shape of macromolecules.

The significance of molecular flexibility manifests itself in a number of ways in chemistry and biology. In enzyme-substrate or receptor-ligand interactions, conformational freedom influences the adaptability of these pairs of molecules. Further, the availability of flexible substituents near pharmacophores influences the possibility of additional binding outside an active site or receptor, leading to inhibition or antagonism. In chemical reactions, the flexibility governs the availability of reactive sites to reagents or the possibility of intermolecular interactions leading to a chemical change. Physical properties of polymers are strongly influenced by the flexibility of the main chain.

It is easy to list the various chemical and biological events influenced by flexibility, but unfortunately efforts to quantitate this structural attribute have been few. Mann<sup>152</sup> analyzed the conformation of alkanes by modifying the number of *gauche* arrangements with Pitzer's steric partition function. Luisi<sup>151</sup> ranked alkanes on a scale of conformational rigidity based on three-states rotational isomerism. Unfortunately, these schemes are designed for acyclic hydrocarbons and have no inherent capability to be adapted to heteroatomic molecules.

An alternative approach to a scheme for the quantitation of flexibility applicable to any type of molecule begins with a definition based on some aspects of molecular structure. Accordingly, we elect to define molecular flexibility on the basis of structural attributes that mitigate against a value ascribed to an infinitely flexible molecule.

A perfectly flexible molecule is assumed to be an endless chain of C(sp<sup>3</sup>) atoms. The flexibility index is infinite. The structural features mitigating against this condition are (1) fewer atoms, (2) the presence of cycles, (3) the presence of branching, and (4) the presence of atoms with covalent radii smaller than C(sp<sup>3</sup>). By quantitating each of these factors, we can arrive at a consistent method for comparing the flexibility of two molecules.

Each of the structural attributes in the definition above has been incorporated into our analysis of shape using the kappa indexes.<sup>59-61</sup> In particular, the count of atoms and the relative cyclicity of molecules have been shown to be

**Table 7** The Flexibility Index  $\phi$  Values for Selected Graphs

Graph	$\phi$
	5.000
	6.000
	3.200
	2.222
	1.543
	1.506
	0.913

encoded into the  ${}^1\kappa$  index. The branching or relative spatial density of molecules is encoded into the  ${}^2\kappa$  index. Finally the presence of non-C(sp<sup>3</sup>) atoms is encoded into the alpha value modifying each kappa index.

Our definition can then be formalized into an expression reflecting the compounding effect of these factors. An index called  $\phi$  is defined:

$$\phi = {}^1\kappa_\alpha {}^2\kappa_\alpha / A \quad [66]$$

Normalization with the count of the number of atoms,  $A$ , permits a comparison of isomers. Information encoded in  $\phi$  can be observed in Table 7. For normal alkanes, the value of  $\phi$  equals the count of the number of bonds in the carbon skeleton. With homologation, the value increases, consistent with our definition. With increased branching or cyclicity, the value decreases as expected. The combined effects of branching and cyclicity leads to a sharp decline in the  $\phi$  value. Finally the effect of unsaturation and heteroatoms results in a decrease in value when these atoms have a covalent radius less than C(sp<sup>3</sup>).

The direct reflection of flexibility in a physical property is not clearly demonstrable. Nevertheless, this attribute clearly plays a role in intermolecular phenomena. It will certainly be of value in reckoning the fit of relatively flexible molecules at enzyme and receptor sites.

---

## SPECIFIC APPLICATION OF KAPPA INDEXES

It is instructive to review briefly a few of the structure analyses in which the kappa shape indexes have played a meaningful role.

### The Pitzer Acentric Factor

Pitzer derived a term useful in corresponding state predictions.<sup>152</sup> The acentric factor  $\omega$  is defined in terms of vapor pressure and is designed to account for the nonideal behavior of gases. In essence, it encodes information about the nonspherical shape of molecules. Using published<sup>154</sup> values of  $\omega$ , Kier has shown an excellent correlation with  $^1\kappa$  and  $^2\kappa$ .<sup>60</sup>

$$\omega = 0.036 \ ^1\kappa + 0.011 \ ^2\kappa + 0.026 \quad [67]$$

$$r = 0.996, s = 0.011, F = 965, n = 20$$

The equation predicts the acentric factor of 20 hydrocarbons within 3% error.

### Comparison with the Taft Steric Parameter

It is useful to compare graph-based shape indexes and empirical measures such as those contributed by Taft.<sup>135</sup> This requires the comparison of kappa shape indexes for molecular groups with values deduced by Taft from hydrolysis rate data.

To calculate kappa values for chemical groups requires dealing with an incomplete molecule, and identifying the attacking atom in some specific manner. Further, there is a possible algebraic problem dealing with small path counts. One proposed solution took advantage of the perceived additivity found for kappa values.<sup>154</sup> The kappa values for an isopropyl group would have no meaning because the attachment atom, the *ipso* atom, is not specifically identified. According to this scheme, the group used was a dimer formed by combining two identical fragments at the *ipso* atoms. Thus, an ethyl group becomes butane, an isopropyl group becomes 2,3-dimethylbutane, and so on. The kappa indexes are calculated for the dimer. To obtain kappa values for the group, dimer values are divided by 2.

A correlation was found between the Taft  $E_s$  values and three kappa indexes:

$$-E_s = -0.039 \ ^0\kappa + 0.78 \ ^1\kappa - 0.34 \ ^3\kappa - 0.63 \quad [68]$$

$$r = 0.9661, s = 0.287, F = 167, n = 46$$

The occurrence of  ${}^0\kappa$  in the equation is illuminating because it suggests a role for symmetry in the steric effect.

In a later paper, Kier<sup>155</sup> expressed the equation above as a new graph-based index capable of predicting steric effects of any group irrespective of its chemical role in the Taft model. This index,  $\Xi$ , is calculated as

$$\Xi = {}^2\kappa_a - {}^0\kappa - {}^3\kappa_a \quad [69]$$

It is anticipated that the  $\Xi$  index can be used in QSAR equations in the same manner as the  $E_s$ , but that no experimental measures are required.

## Enzyme Inhibitors

Shape quantitation is expected to play a useful role in enzyme-inhibitor studies because of the necessity of shape complementarity. Quantitation of the shape of enzyme inhibitor candidate molecules has revealed some good correlations of potential utility in designing more effective compounds. These include a series of aminoalkyl adenylates binding to isoleucine-RNA synthetase,<sup>60</sup> alcohols inhibiting aniline hydroxylation,<sup>60</sup> hydroxy quinolines inhibiting maleate dehydrogenase,<sup>61</sup> and the effect of benzimidazoles in displacing the dihydro-safrole metabolite from a complex with cytochrome P-450.<sup>156</sup>

## Toxicity Analysis

A large group of substituted benzenes have been analyzed for their toxicities against fathead minnows. One relationship with structure revealed a prominent role of shape in addition to topological factors. An equation was developed in which the measured toxicity, given as  $-\log IC_{50}$ , related to both chi and kappa indexes and an indicator variable I, which describes the presence of two nitro groups, ortho or para to each other.<sup>157</sup> The equation and statistics are:

$$-\log IC_{50} = 1.430 {}^1\kappa_a - 2.713 {}^1\chi - 1.036 {}^6\chi_P^v + 1.10 I + 2.90 \quad [70]$$

$r = 0.940, s = 0.26, F = 1114, n = 65$

This large set of molecules presents a very good opportunity to explore the role of certain shape attributes in the mechanism of toxicity as well as the other factors represented by the chi indexes. The study also is useful in creating a general model for the prediction of toxicities of related benzene derivatives.

The kappa shape indexes are relatively new descriptors, and few studies have been reported to date. It is expected that the kappa indexes will generally be used along with other topological indexes in QSAR equations rather than by themselves as in most of the examples above. Shape is not the sole determinant

of activities or properties in most cases. There are electronic aspects of structure to be considered along with other topological and size effects. As the kappa indexes are considered by QSAR investigators, it is anticipated that they will find use in a variety of studies.

---

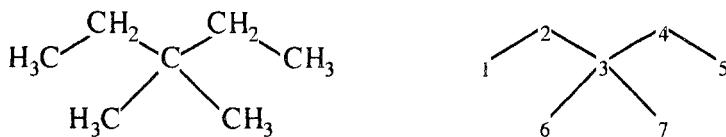
## CHARACTERIZATION OF SKELETAL ATOMS, THE TOPOLOGICAL STATE

Based on the use of chemical graph theory as described above, various indexes of molecular structure have been developed. These indexes may all be termed topological indexes. In the molecular connectivity method, indexes have been developed to characterize various aspects of molecular structure. The kappa shape indexes were developed so that shape measures could be directly entered in QSAR analyses. Each of these indexes characterizes the whole molecule with respect to one or more aspects of structure. In chemistry it is also of interest to characterize the skeletal atoms. In this final section we review briefly an investigation of the skeletal atoms as vertexes in the molecular graph as a basis for an atom descriptor.

Investigation of an index for atom characterization is not new. In an early study, Hall and Kier<sup>158</sup> investigated the relation between atomic environment of atoms in alkanes and the net atomic charge as computed by a semiempirical MO method. Using quantities computed from one, two, and three atom paths, they found high correlation to the partial charge. These path dependent graph quantities also reflected the topological equivalence of the skeletal atoms. Later, Kier used a related approach as a basis for determination of topological equivalence so that skeletal atoms could be classified into topologically equivalent groups. This classification scheme was the basis for computation of an information-based graphical index that was used in QSAR equations.<sup>154</sup>

As a further development, Hall and Kier introduced the idea of the topological state of the graph vertex (which represents the skeletal atom).<sup>62</sup> The topological state of the skeletal atom is the basis of both the determination of topological equivalence as well as computation of graphical indexes which can be related to the properties of the atoms in the molecule. The atom is viewed in the context of the full topology of the whole molecule.

Based on the complete set of paths from a given atom to every other atom in the whole molecule, the topological state encodes information on the entire topological environment of the given atom. The set of all paths terminating on a given atom may be considered as a complete characterization of that atom in the molecule. Consider, for example, the molecule 3,3-dimethylpentane as shown in Figure 4. The following list shows all the paths that terminate on atom 1. Each path is given along with the atom identification numbers of the two ends of the path.



**Figure 4** The structural formula for 3,3-dimethylpentane and the numbered hydrogen-suppressed graph.

1:1	$\text{CH}_3$	(5:5)
1:2	$\text{CH}_3-\text{CH}_2$	(5:4)
1:3	$\text{CH}_3-\text{CH}_2-\text{C}$	(5:3)
1:4	$\text{CH}_3-\text{CH}_2-\text{C}-\text{CH}_2$	(5:2)
1:5	$\text{CH}_3-\text{CH}_2-\text{C}-\text{CH}_2-\text{CH}_3$	(5:1)
1:6	$\text{CH}_3-\text{CH}_2-\text{C}-\text{CH}_3$	(5:6)
1:7	$\text{CH}_3-\text{CH}_2-\text{C}-\text{CH}_3$	(5:7)

This collection of path information characterizes atom 1 within the molecular skeleton, in the topological sense, using character string information for the atomic symbols. Examination of the graph shows that exactly the same set of vertex sequences applies to vertex 5. The corresponding set of vertex pairs for 5 is given in parentheses following each vertex sequence. In this way, it is established that vertexes 1 and 5 are topologically equivalent. It can be shown by inspection that no other vertex has the same set of vertex sequences.

In the spirit of chemical graph theory, a numerical expression was developed to encode the structure information represented in the paths. The numerical value for the path from vertex  $i$  to vertex  $j$  is the entry  $t_{ij}$  in a topological state matrix  $T$ . The topological state value,  $T_i$ , for vertex  $i$  is, then, the sum of the entries for row  $i$ .

$$T_i = \sum_{j=1}^A t_{ij} \quad [71]$$

The key to useful topological state values is an appropriate form for the  $t_{ij}$  values. Hall and Kier have shown that simple forms, such as the graph distance  $d_{ij}$ , are not useful because they fail to indicate proper topological equivalence.<sup>62</sup> To ensure representation of topological equivalence, two features of the paths must be encoded: (1) atomic identity and (2) the sequence of atoms in each path. It has been shown that both these characteristics can be encoded as follows. Atomic identity can be encoded using the molecular connectivity valence delta value,  $\delta^v$ . The discussions concerning chi indexes and related quantities have shown the validity of the valence delta value as a characterization of atoms.

It is not as easy to preserve the atom sequence in the paths. However, Hall and Kier<sup>62</sup> have shown that topological equivalence can be preserved when the

topological state is made a function of the path distance and have presented a set of relationships for that purpose. The topological state quantity  $t_{ij}$  is taken as a product of two functions: the geometric mean (GM) of the valence delta values of the vertexes in the path and a simple function of the length of the path, counted as the number of atoms in the path,  $d$ .

$$t_{ij} = [\text{GM}_{ij}]^a f(d) = \left[ \prod_{k=1}^{n_{ij}} \delta_k^v \right]^{a/n_{ij}} f(d) \quad [72]$$

The geometric mean is taken either directly or as the reciprocal;  $a = \pm 1$ . Several functions of  $d$  have been considered, such as powers of  $d$ ;

$$f(d) = d^b \quad [73]$$

where  $b = 0, \pm 1, \pm 2, \pm 3$ , etc. More complex functions can, of course, be used, such as  $u + vd + wd^2 + \dots$ , where  $u, v, w$  are simple constants such as  $+1$ . The expressions investigated by Hall and Kier all take the simple form

$$t_{ij} = [\text{GM}_{ij}]^a d^b \quad [74]$$

This expression is a graph invariant generator. As will be shown later, this topological state expression will be used to develop a family of indexes for the whole graph that is a highly discriminating index of molecular graphs.

Examples of the topological matrix for such molecules as 3-ethyl-3-methylhexane and 1-isopropyl-2,3-dimethylcyclopropane have been given.<sup>62</sup> All computations of the topological state are carried out by the program MOLCONN-X.<sup>159</sup>

This topological formalism has been evaluated for determination of topological equivalence for a large number of molecules, including the 308 alkanes from ethane through the undecanes, 30 polycyclic aromatic hydrocarbons, and all the polymethyl- and polyethylbenzenes. A large number of unusual graphs have been suggested by Randić<sup>160,161</sup> and others,<sup>162</sup> and many more graphs of various sizes and types, including those with many heteroatoms, have been studied. There have been no failures in proper topological determination.<sup>62</sup>

Another index in the topological state formalism has been developed. The total topological index  $\tau$ , has been defined, based on a summation of the topological state  $T_i$  values.<sup>62</sup>

$$\tau = \sum_{i=1}^A t_i + \sum_{i=1}^A \sum_{j>i} t_{ij} \quad [75]$$

This index has been found to be unique for all the graphs that have so far been examined, including all those mentioned above that were examined for correct prediction of topological equivalence. At this time it is suggested that the  $\tau$  index

<u>Heptane Graph</u>	<u>Total Topological Index <math>\tau</math></u>
	16.540
	16.846
	17.005
	17.162
	17.176
	17.407
	17.422
	17.651
	17.891

Figure 5 Heptane isomers ranked according to the  $\tau$  index, indicating that one of the characteristics of the  $\tau$  index is molecular compactness. The topological state algorithm used for this example is  $t_{ij} = GM_{ij}/d_{ij}^2$ .

is the most highly discriminating topological index yet developed. The extent of its discrimination power is being investigated. Examples of  $\tau$  values of heptane isomers are given in Figure 5. Because the  $\tau$  index is highly discriminating, it may be useful as a QSAR correlate. Hall and Kier<sup>62</sup> reported two QSAR examples using the total topological index  $\tau$ . One of these will be given here. Berger et al.<sup>163</sup> reported the inhibition of *T. mentagrophytes* by a set of ethers of glycerol and trimethyleneglycerol. The 28 alkyl- and chloro-substituted compounds were analyzed using chi indexes and the  $\tau$  index. The activity is the negative logarithm of the concentration which gives 50% inhibition,  $pC_{50}$ :

$$pC_{50} = 4.101 \cdot {}^3\chi_P^\nu - 4.510 \cdot {}^3\chi_P + 0.1681 \tau + 1.188 \quad [76]$$

$$r = 0.963, s = 0.14, F = 103, n = 28$$

The average relative error is 3.9%, and no observation has a relative error greater than 10%. These results are significantly better than those given by the Hansch model. Hansch and Lien<sup>164</sup> found it necessary to delete two observations to achieve  $r = 0.911$ ,  $s = 0.22$ . When the full data set is used, the statistics are even worse:  $r = 0.878$ ,  $s = 0.24$ . The potential value of this total topological index  $\tau$  is yet to be explored. Because it is a very highly discriminating index, there may be a useful role for it in QSAR.

This review gives clear evidence of the QSAR versatility afforded by these topological indexes. Much work remains in the development of strategies for optimum application of these indexes, including implementation of statistical techniques and introduction of methods for transformation to orthogonal sets of indexes. Further, this topological index approach does have limitations. For properties highly dependent on very localized regions of molecules, whole molecule indexes may not be satisfactory. In addition, properties that arise largely from very specific electronic effects may not be well treated, including charge-transfer, covalent bond formation, or electronically directed effects not related to ground state properties. One effort to include electronic and topological attributes has recently been reported.<sup>63-66</sup> The electropotential state index encodes such information in an index value for each atom.<sup>67</sup> It is our hope that continued efforts will more fully equip chemists to describe structure–property relationships.

---

## REFERENCES

1. A. T. Balaban, Ed., *Chemical Applications of Graph Theory*. Academic Press, New York, 1976.
2. L. B. Kier and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*. Academic Press, New York, 1976.
3. L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure-Activity Analysis*. Research Studies Press, Wiley, Letchworth, England, 1986.
4. N. Trinajstić, *Chemical Graph Theory*, Vols. I, II. CRC Press, Boca Raton, FL, 1983.
5. R. B. King, Ed., *Chemical Applications of Topology and Graph Theory*. Elsevier, Amsterdam, 1983.
6. D. H. Rouvray, *Am. Sci.*, **61**, 729 (1973). The Search for Useful Topological Indices in Chemistry.
7. D. H. Rouvray, *Sci. Am.*, **255**, 40 (1986). Predicting Chemistry from Topology.
8. P. J. Hansen and P. C. Jurs, *J. Chem. Educ.*, **65**, 574 (1988). Chemical Applications of Graph Theory II: Isomer Enumeration.
9. A. Sablić and N. Trinajstić, *Acta Pharm. Jugosl.*, **31**, 189 (1981). Quantitative Structure-Activity Relationships: The Role of Topological Indices.
10. M. Randić, *J. Am. Chem. Soc.*, **97**, 6609 (1975). On the Characterization of Molecular Branching.
11. L. B. Kier and L. H. Hall, *J. Pharm. Sci.*, **65**, 1806 (1976). Molecular Connectivity. VII. Specific Treatment of Heteroatoms.
12. L. H. Hall, L. B. Kier, and W. J. Murray, *J. Pharm. Sci.*, **64**, 1974 (1975). Molecular Connectivity. II. Relationship to Water Solubility and Boiling Point.

13. L. B. Kier, W. J. Murray, M. Randić, and L. H. Hall, *J. Pharm. Sci.*, **65**, 1226 (1976). Molecular Connectivity. V. Connectivity Series Concept Applied to Density.
14. W. J. Murray, L. H. Hall, and L. B. Kier, *J. Pharm. Sci.*, **64**, 1978 (1975). Molecular Connectivity. III. Relationship to Partition Coefficients.
15. M. J. M. Wells, R. C. Clark, and R. M. Patterson, *J. Chromatogr. Sci.*, **19**, 573 (1981). Correlation of Reserved-Phase Capacity Factors for Barbiturates with Biological Activities, Partition Coefficients, and Molecular Connectivity Indexes.
16. J. C. Boyd, J. S. Millership, and A. D. Woolfson, *J. Pharm. Pharmacol.*, **34**, 364 (1982). The Relationship between Molecular Connectivity and Partition Coefficients.
17. T. R. McGregor, *J. Chromatogr. Sci.*, **17**, 314 (1979). Connectivity Parameters as Predictors of Retention in Gas Chromatography.
18. R. Kaliszan and H. Lamparczyk, *J. Chromatogr. Sci.*, **16**, 246 (1978). A Relationship Between the Connectivity Indices and Retention Indices of Polycyclic Aromatic Hydrocarbons.
19. F. Calixto and A. Raso, *Chromatographia*, **15**, 521 (1982). Retention Index, Connectivity Index and van der Waals Volume of Alkanes.
20. M. J. M. Wells, R. C. Clark, and R. M. Patterson, *J. Chromatogr.*, **235**, 61 (1982). Investigation of N-Alkylbenzamides by Reversed-Phase Liquid Chromatography. III. Correlation of Chromatographic Parameters with Molecular Connectivity Indices for C1-C5 N-Alkylbenzamides.
21. L. Buydens, D. L. Massart, and P. Geerlings, *Anal. Chem.*, **55**, 738 (1983). Prediction of Gas Chromatographic Retention Indexes with Topological, Physicochemical, and Quantum Chemical Parameters.
22. G. Szasz, O. Papp, J. Vamos, K. Hanko-Novak, and L. B. Kier, *J. Chromatogr.*, **269**, 144 (1984). Relationships Between Molecular Connectivity Indexes, Partition Coefficients and Chromatographic Parameters.
23. P. J. Doherty, R. M. Hoes, and A. Robbat, *Anal. Chem.*, **56**, 2697 (1984). Relation Between Gas Chromatographic Retention Indices and Molecular Connectivities of Nitrated Polycyclic Aromatic Hydrocarbons.
24. A. Sabljic, *J. Chromatogr.*, **314**, 1 (1984); **319**, 1 (1985). Calculation of Retention Indices by Molecular Topology: Chlorinated Benzenes.
25. T. DiPaolo, L. B. Kier, and L. H. Hall, *Mol. Pharmacol.*, **13**, 31 (1977). Molecular Connectivity and Structure-Activity Relationship of General Anesthetics.
26. T. Dipaolo, *J. Pharm. Sci.*, **67**, 564 (1978). Structure-Activity Relationships of Anesthetic Ethers Using Molecular Connectivity.
27. T. Dipaolo, *J. Pharm. Sci.*, **67**, 566 (1978). Molecular Connectivity in Quantitative Structure-Activity Relationship Study of Anesthetic and Toxic Activity of Aliphatic Hydrocarbons, Ethers, and Ketones.
28. T. Dipaolo, *J. Pharm. Sci.*, **68**, 39 (1979). Molecular Connectivity Study of Halocarbon Anesthetics.
29. M. C. Bindal, P. Singh, and S. P. Gupta, *Arzneim.-Forsch.*, **30**, 234 (1980). Quantitative Correlation of Anesthetic Potencies of Halogenated Hydrocarbons with Boiling Point and Molecular Connectivity.
30. L. B. Kier and L. H. Hall, *J. Med. Chem.*, **20**, 1631 (1977). Structure-Activity Studies on Hallucinogenic Amphetamines Using Molecular Connectivity.
31. L. B. Kier and R. A. Glennon, *Life Sci.*, **22**, 1589 (1978). Psychotomimetic Phenalkylamines as Serotonin Antagonists: An SAR Analysis.
32. R. A. Glennon and L. B. Kier, *Eur. J. Med. Chem.*, **13**, 219 (1978). LSD Analogs as Serotonin Antagonists: A Molecular Connectivity SAR Analysis.
33. R. A. Glennon, L. B. Kier, and A. T. Shulgin, *J. Pharm. Sci.*, **68**, 906 (1979). Molecular Connectivity Analysis of Hallucinogenic Mescaline Analogs.
34. A. J. Richard and L. B. Kier, *J. Pharm. Sci.*, **69**, 124 (1980). SAR Analysis of Hydrazide Monoamine Oxidase Inhibitors Using Molecular Connectivity.

35. A. Sabljic and M. Protic-Sabljic, *Mol. Pharmacol.*, **23**, 213 (1983). Quantitative Structure-Activity Study on the Mechanism of Inhibition of Microsomal *p*-Hydroxylation of Aniline by Alcohols.
36. L. B. Kier, T. DiPaolo, and L. H. Hall, *J. Theor. Biol.*, **67**, 585 (1977). Structure-Activity Studies on Odor Molecules Using Molecular Connectivity.
37. R. J. Gardner, *J. Sci. Food Agric.*, **31**, 23 (1980). Correlation of Bitterness Thresholds of Amino Acids and Peptides with Molecular Connectivity.
38. W. J. Spillane, G. McGlinchey, I. O. Muricheartaigh, and G. A. Benson, *J. Pharm. Sci.*, **72**, 934 (1983). Structure-Activity Studies on Sulfamate Sweeteners. 3. Structure-Taste Relationships for Heterosulfamates.
39. L. B. Kier, *J. Pharm. Sci.*, **69**, 416 (1980). Molecular Structure Influencing Either a Sweet or Bitter Taste Among Aldoximes.
40. A. Sabljic and M. Protic, *Chem. Biol. Interact.*, **42**, 301 (1984). Molecular Connectivity: A Novel Method for Prediction of Bioconcentration Factor of Hazardous Chemicals.
41. A. Sabljic and M. Protic, *Bull. Environ. Contam. Toxicol.*, **28**, 162 (1982). Relationship Between Molecular Connectivity Indices and Soil Sorption Coefficients of Polycyclic Aromatic Hydrocarbons.
42. L. H. Hall and L. B. Kier, *J. Pharm. Sci.*, **67**, 1743 (1978). Molecular Connectivity and Substructure Analysis.
43. L. H. Hall and L. B. Kier, *Eur. J. Med. Chem.*, **13**, 89 (1978). A Comparative Analysis of Molecular Connectivity, Hansch, Free-Wilson and Darc-Pelco Methods in the SAR of Halogenated Phenols.
44. A. K. Samata, S. Ray, S. C. Basak, and S. K. Bose, *Arzneim.-Forsch.*, **32**, 1515 (1982). Molecular Connectivity and Antifungal Activity.
45. R. Koch, *Chemosphere*, **11**, 925 (1982). Molecular Connectivity and Acute Toxicity of Environmental Pollutants.
46. W. T. Schultz, L. B. Kier, and L. H. Hall, *Bull. Environ. Contam. Toxicol.*, **28**, 373 (1982). Structure-Toxicity Relationships of Selected Nitrogenous Heterocyclic Compounds. III. Relations Using Molecular Connectivity.
47. A. Sabljic, *Bull. Environ. Contam. Tox.*, **30**, 80 (1983). Quantitative Structure-Toxicity Relationship of Chlorinated Compounds: A Molecular Connectivity Investigation.
48. L. H. Hall and L. B. Kier, *Bull. Environ. Contam. Tox.*, **32**, 354 (1984). Molecular Connectivity of Phenols and Their Toxicity to Fish.
49. H. Govers, C. Ruepert, and H. Aiking, *Chemosphere*, **13**, 227 (1984). Quantitative Structure-Activity Relationships for Polycyclic Aromatic Hydrocarbons: Correlation Between Molecular Connectivity, Physicochemical Properties, Bioconcentration and Toxicity in *Daphnia pulex*.
50. L. B. Kier, R. J. Simons, and L. H. Hall, *J. Pharm. Sci.*, **67**, 725 (1978). Structure-Activity Studies on Mutagenicity of Nitrosamines Using Molecular Connectivity.
51. P. C. Jurs, J. T. Chou, and M. Yuan, *J. Med. Chem.*, **22**, 476 (1979). Computer-Assisted Structure-Activity Studies of Chemical Carcinogens. A Heterogeneous Data Set.
52. M. Yuan and P. C. Jurs, *Tox. Appl. Pharmacol.*, **52**, 294 (1980). Computer-Assisted Structure-Activity Studies of Chemical Carcinogens: A Heterogeneous Data Set.
53. Z. Melkova, *Cesk. Farm.*, **33**, 107 (1984). Utilization of the Index of Molecular Connectivity in the Study of Antitumor Activity of a Group of Benzo(c)fluorene Derivatives.
54. A. J. Stuper and P. C. Jurs, *J. Pharm. Sci.*, **67**, 745 (1978). Structure-Activity Studies of Barbiturates using Pattern Recognition Techniques.
55. D. R. Henry and J. H. Block, *J. Med. Chem.*, **22**, 465 (1979). Classification of Drugs by Discriminant Analysis Using Molecular Connectivity Values.
56. D. R. Henry and J. H. Block, *Eur. J. Med. Chem.*, **15**, 133 (1980). Steroid Classification by Discriminant Analysis Using Fragment Molecular Connectivity.

57. D. R. Henry and J. H. Block, *J. Pharm. Sci.*, **69**, 1030 (1980). Pattern Recognition of Steroids Using Fragment Molecular Connectivity.
58. Y. Takahashi, Y. Miashita, and Y. Tanaka, *J. Pharm. Sci.*, **73**, 737 (1985). Discriminative Structural Analysis Using Pattern Recognition Techniques in the Structure-Taste Problem of Peril-lartines.
59. L. B. Kier, *Quant. Struct.-Act. Relat.*, **4**, 109 (1985). A Shape Index from Molecular Graphs.
60. L. B. Kier, *Quant. Struct.-Act. Relat.*, **5**, 1, 7 (1985). Distinguishing Atom Differences in a Molecular Graph Shape Index. [Covalent radii are from V. Schomaker and D. P. Stevenson, *J. Am. Chem. Soc.*, **63**, 37 (1941). Some Revisions of the Covalent Radii and the Additivity Rule for the Lengths of Partially Ionic Single Covalent Bonds.]
61. L. B. Kier, *Acta Pharm. Jugosl.*, **36**, 171 (1986). Indexes of Molecular Shape from Chemical Graphs.
62. L. H. Hall and L. B. Kier, *Quant. Struct.-Act. Relat.*, **9**, 115 (1990). Determination of Topological Equivalence in Molecular Graphs from the Topological State.
63. L. B. Kier and L. H. Hall, *Pharmaceut. Res.*, **7**, 801 (1990). An Electrotopological State Index for Atoms in Molecules.
64. L. B. Kier and L. H. Hall, *J. Math. Chem.*, in press (1991). An Index of Electrotopological State of Atoms in Molecules.
65. L. H. Hall, B. K. Mohney, and L. B. Kier, *J. Chem. Inf. Comput. Sci.*, **31**, 76 (1991). The Electrotopological State: Structure Information at the Atomic Level for Molecular Graphs.
66. L. H. Hall, B. K. Mohney, and L. B. Kier, *Quant. Struct.-Act. Relat.*, **10**, 43 (1991). The Electrotopological State: An Atom Index for QSAR.
67. L. B. Kier and L. H. Hall, in *Drug Design*. B. Testa, Ed. An Atom-Centered Index for Drug QSAR Models. 1991, in press.
68. E. L. Eliel, *Stereochemistry of Carbon Compounds*. McGraw-Hill, New York, 1962, p. 1.
69. O. Ore, *The Theory of Graphs*. Methuen, London, 1962.
70. R. G. Busacker and T. L. Saaty, *Finite Graphs and Networks*. McGraw-Hill, New York, 1965.
71. R. Wilson, *Introduction to Graph Theory*. Academic Press, New York, 1972.
72. F. Harary, Ed., *New Directions in Graph Theory*. Academic Press, New York, 1973.
73. H. Wiener, *J. Am. Chem. Soc.*, **69**, 17, 2636 (1947). Structural Determination of Paraffin Boiling Points.
74. H. Wiener, *J. Phys. Chem.*, **52**, 425, 1082 (1948). Vapor Pressure-Temperature Relationships among the Branched Paraffin Hydrocarbons.
75. J. R. Platt, *J. Phys. Chem.*, **56**, 328 (1952). Prediction of Isomeric Differences in Paraffin Properties.
76. J. R. Platt, in *Encyclopedia of Physics*, Vol. 37. S. Flügge, Ed. Springer-Verlag, Berlin, 1961, p. 173.
77. E. Cayley, *Phil. Mag.*, **67**, 444 (1874). On the Mathematical Theory of Isomers.
78. H. Schiff, *Chem. Ber.*, **8**, 1542 (1875). Zur Statistik Chemischer Verbindungen.
79. H. R. Henze and C. M. Blair, *J. Am. Chem. Soc.*, **56**, 157 (1934). The Number of Structural Isomers of the More Important Types of Aliphatic Compounds.
80. K. Altenburg, *Kolloid-Z.*, **178**, 112 (1961). Zur Berechnung des Radius Verweigter Moleküle.
81. K. Altenburg, *Brennst.-Chem.*, **47**, 100, 331 (1966). Dependence of the Boiling Point of Isomeric Hydrocarbons on the Form of the Molecules. I. Theoretical Basis and Applications to Isomeric Acyclic Alkanes.
82. M. Gordon and G. R. Scantlebury, *Trans. Faraday Soc.*, **60**, 605, (1964). Theory of Ring-chain Equilibrium in Branched Non-Random Polycondensation Systems.

83. H. Hosoya, *Bull. Chem. Soc. Jpn.*, **44**, 2332 (1971). Topological Index. A Newly Proposed Quantity Characterizing the Topological Nature of Structural Isomers of Saturated Hydrocarbons.
84. E. A. Smolenskii, *J. Am. Chem. Soc.*, **97**, 6609 (1975). Application of the Theory of Graphs to Calculations of the Additive Structural Properties of Hydrocarbons.
85. S. Glasstone, *Textbook of Physical Chemistry*. 2nd ed., Van Nostrand Reinhold, Princeton, NJ, 1946, Chap. 8.
86. J. Sugden, *J. Chem. Soc.*, **125**, 32 (1924). The Variation of Surface Tension with Temperature and Some Related Functions.
87. A. L. von Steiger, *Chem. Ber.*, **54**, 1381 (1921). Principle of Constant Sum of Energies of Formation and Their Distribution on the Unions in Aromatic Substances.
88. K. Fajans and J. Knorr, *Chem. Ber.*, **59**, 249 (1926). Refractive Index of Organic Substances and Atomic Structure.
89. C. P. Smyth, *Phil. Mag.*, **50**, 361, 715 (1925). Electric Moment and Molecular Structure.
90. L. Pauling, *The Nature of the Chemical Bond*, 3rd ed. Cornell University Press, Ithaca, NY, 1960, Chap. 3.
91. K. Fajans, *Chem. Ber.*, **53**, 643 (1920). The Energy of the Atomic Unions in the Diamond and in Aliphatic Hydrocarbons.
92. F. Klages, *Chem. Ber.*, **82**, 358 (1949). Über eine Verbesserung der additiven Berechnung von Verbrennungswärmen und der Berechnung der Mesomeric Energie aus Verbrennungswärmen.
93. G. S. Parks and H. M. Huffman, *Free Energies of Some Organic Compounds*. ACS Monograph No. 60, Chemical Catalog Co., New York, 1932.
94. S. W. Benson and J. H. Buss, *J. Chem. Phys.*, **29**, 546 (1958). Additivity Rules for the Estimation of Molecular Properties. Thermodynamic Properties.
95. V. M. Tatevskii, V. A. Benderskii, and S. S. Yarovsi, *Rules and Methods for Calculating Physicochemical Properties of Paraffinic Hydrocarbons*. Pergamon Press, Oxford, 1961.
96. J. L. Franklin, *Ind. Eng. Chem.*, **41**, 1070 (1949). Prediction of Heat and Free Energies of Organic Compounds.
97. G. R. Somayajulu and B. J. Zwolinski, *Trans. Faraday Soc.*, **62**, 2327 (1966). Generalized Treatment of Alkanes. Part 1.
98. G. R. Somayajulu and B. J. Zwolinski, *Trans. Faraday Soc.*, **68**, 1971 (1972). Generalized Treatment of Alkanes. Part 2.
99. R. B. Woodward, *J. Am. Chem. Soc.*, **63**, 1123 (1941). Structure and the Absorption Spectra of  $\alpha,\beta$  Unsaturated Ketones.
100. D. Bonchev, *Information Theoretic Indices for Characterization of Chemical Structures*. Research Studies Press, Wiley, Letchworth, England, 1975.
101. L. B. Kier and L. H. Hall, *J. Pharm. Sci.*, **70**, 583 (1981). Derivation and Significance of Valence Molecular Connectivity.
102. L. B. Kier and L. H. Hall, *J. Pharm. Sci.*, **72**, 1170 (1983). A General Definition of Valence Delta-Values for Molecular Connectivity.
103. A. Bondi, *J. Phys. Chem.*, **68**, 441 (1964). van der Waals Volumes and Radii.
104. R. S. Mulliken, *J. Chem. Phys.*, **2**, 782 (1934). New Electronegativity Scale; Together with Data on Valence States and on Valence State Ionization Potentials and Electron Affinities.
105. J. Hinze and H. H. Jaffe, *J. Am. Chem. Soc.*, **84**, 540 (1962). Electronegativity. I. Orbital Electronegativity of Neutral Atoms.
106. L. B. Kier and L. H. Hall, *Quant. Struct.-Act. Relat.*, **2**, 163 (1983). Estimation of Substituent Group Electronic Influence from Molecular Connectivity Delta Values.
107. P. Seybold, M. A. May, and M. L. Gargas, *Acta Pharm. Jugosl.*, **36**, 253 (1986). On the Physical Properties of Halogenated Hydrocarbons.

108. P. G. Seybold, M. May, and U. A. Bagal, *J. Chem. Educ.*, **64**, 575 (1987). Molecular Structure-Property Relationships.
109. D. H. Rouvray, *Acta Pharm. Jugosl.*, **36**, 239 (1986). The Prediction of Biological Activity Using Molecular Connectivity Indexes.
110. A. Sabljić, *J. Chromatogr.*, **319**, 1 (1985). Calculation of Reduction Indices by Molecular Topology. Chlorinated Benzenes.
111. M. L. Lee, D. L. Vassilaros, C. M. White, and M. Novotny, *Anal. Chem.*, **51**, 768 (1979). Retention Indices for Programmed-Temperature Capillary-Column Gas Chromatography of Polycyclic Aromatic Hydrocarbons.
112. R. Kaliszan and H. Foks, *Chromatographia*, **10**, 346 (1977). The Relationship between the  $R_m$  Values and the Connectivity Indices for Pyrazine Carbothioamide Derivatives.
113. L. H. Hall, unpublished data.
114. G. I. Spivakovskii, A. I. Tischenko, I. I. Zaslovskii, and N. S. Wulfson, *J. Chromatogr.*, **144**, 1, 1977. Calculation of Retention Indices of Compounds from Their Structural Formulae for Combined Identification by Gas Chromatography-Mass Spectrometry. Application to Aliphatic Alcohols and Saturated Hydrocarbons.
115. A. Robbat, N. P. Curso, P. J. Doherty, and D. Marshall, *Anal. Chem.*, **58**, 2072 (1986). Multivariate Relationships Between Gas Chromatographic Retention Index and Molecular Connectivity of Mononitrated Polycyclic Aromatic Hydrocarbons.
116. A. Robbat, N. P. Curso, P. J. Doherty, and M. Wolf, *Anal. Chem.*, **58**, 2078 (1986). Gas Chromatographic Chemiluminescent Detection and Evaluation of Predictive Models for Identifying Nitrated Polycyclic Aromatic Hydrocarbons in a Diesel Fuel Particulate Extract.
117. A. Robbat, G. Xyrafas, and D. Marshall, *Anal. Chem.*, **60**, 2982 (1988). Prediction of Gas Chromatographic Retention Characteristic of Polychlorinated Biphenyls.
118. J. T. Edward, P. G. Farrell, and F. Shahidi, *J. Phys. Chem.*, **82**, 2310 (1978). Partial Molal Volumes of Organic Compounds in Carbon Tetrachloride. I. Alkanes. Conformational Effects.
119. J. D. Cox and G. Pilcher, *Thermochemistry of Organic and Organometallic Compounds*. Academic Press, New York, 1970.
120. K. Watanabe, T. Nakayama, and J. Mottl, *J. Quant. Spectrosc. Radiat. Transfer*, **2**, 369, (1962). Ionization Potentials of Some Molecules.
121. L. H. Hall and D. Aaserud, *Quant. Struct.-Act. Relat.*, **8**, 196 (1989). Structure-Activity Models for Molar Refraction of Alkylsilanes Based on Molecular Connectivity.
122. L. B. Kier and L. H. Hall, *J. Med. Chem.*, **20**, 1631 (1977). Structure-Activity Studies on Hallucinogenic Amphetamines Using Molecular Connectivity.
123. L. B. Kier and L. H. Hall, *J. Pharm. Sci.*, **67**, 1408 (1978). Molecular Connectivity of Muscarinic Receptor Affinity of Acetylcholine Antagonists.
124. A. Sabljić and M. Protić-Sabljić, *Mol. Pharmacol.*, **23**, 213 (1982). Quantitative Structure-Activity Study on the Mechanism of Inhibition of Microsomal *p*-Hydroxylation of Aniline by Alcohols.
125. I. Tamm, K. Folkers, C. H. Shunk, D. Heyl, and F. L. Horofall, *J. Exp. Med.*, **98**, 245 (1953). Inhibition of Influenza Virus Multiplication by Alkyl Derivatives of Benzimidazole.
126. W. van Valkenburg, Ed., *Biological Correlations*. ACS Advances in Chemistry Series 114, American Chemistry Society, Washington, D.C.
127. L. H. Hall and L. B. Kier, *J. Mol. Struct. (THEOCHEM)*, **134**, 309 (1986). Molecular Connectivity and Total Response Surface Optimization.
128. P. O. Mager, *Toxicol. Lett.*, **11**, 67 (1982). Structure-Neurotoxicity Relationships Applied to Organophosphorous Pesticides.
129. SAS Institute Inc., Cary, North Carolina.
130. L. H. Hall and H. D. Stewart, *Quant. Struct.-Act. Relat.*, in press (1991). Total Response Surface Analysis and Molecular Connectivity: Bioconcentration Factors for Halogenated Hydrocarbons.

131. G. W. Adamson and D. Bawden, *J. Chem. Inf. Comput. Sci.*, **16**, 161 (1976). An Empirical Method of Structure-Activity Correlation for Polysubstituted Cyclic Compounds Using Wiswesser Line Notation.
132. G. W. Adamson and J. A. Bush, *J. Chem. Soc. Perkin Trans.*, **1**, 168 (1976). Evaluation of an Empirical Structure-Activity Relationship for Property Prediction in a Structurally Diverse Group of Local Anesthetics.
133. A. Y. Meyer, *Chem. Soc. Rev.*, **15**, 449 (1986). The Size of Molecules.
134. P. G. Mezey, in *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1990, pp. 265–294. Molecular Surfaces.
135. R. W. Taft, in *Steric Effects in Organic Chemistry*. M. S. Newman, Ed., Wiley, New York, 1956, p. 556.
136. M. Charton, *J. Am. Chem. Soc.*, **91**, 615 (1969). The Nature of the Ortho Effect. II. Composition of the Steric Parameters.
137. D. F. DeTar and C. J. Tenpas, *J. Am. Chem. Soc.*, **98**, 7903 (1976). Theoretical Calculation of the Steric Parameters.
138. R. B. Hermann, *J. Phys. Chem.*, **76**, 2754 (1972). Theory of Hydrophobic Bonding. II. Correlation of Hydrocarbon Solubility in Water with Solvent Cavity Surface Area.
139. R. S. Pearlman, in *Physical Chemical Properties of Drugs*. S. H. Yalkowski, Ed., Dekker, New York, 1980, pp. 321–347. Molecular Surface Areas and Volumes and Their Use in Structure-Activity Relationships.
140. A. Verloop, W. Hoogenstraaten, and J. Tipker, in *Drug Design*. Vol. 7, E. J. Ariens, Ed., Academic, New York, 1976, pp. 165–207. Development and Application of New Steric Substituent Parameters in Drug Design.
141. Z. Simon and Z. Szabadai, *Angew. Chem.*, **86**, 802 (1974). Minimal Steric Difference Parameter and the Importance of Steric Fit for Structure-Biological Activity Correlations.
142. A. J. Hopfinger, *J. Am. Chem. Soc.*, **102**, 7196 (1980). A QSAR Investigation of Dihydrofolate Reductase.
143. R. H. Rohrbaugh and P. C. Jurs, *Anal. Chem.*, **59**, 1048 (1987). Molecular Shape and the Prediction of High-Performance Liquid Chromatographic Retention Indexes of Polycyclic Aromatic Hydrocarbons. See also, P. C. Jurs, in *Reviews in Computational Chemistry*, K. B. Lipkowitz and D. B. Boyd, Eds., VCH Publishers, New York, 1990, pp. 169–212. Chemometrics and Multivariate Analysis in Analytical Chemistry.
144. D. Bonchev and N. Trinajstić, *J. Chem. Phys.*, **67**, 4517 (1977). Information Theory, Distance Matrix and Molecular Branching.
145. A. T. Balaban, *Theor. Chem. Acta*, **53**, 355 (1979). Chemical Graphs. XXXIV. Five New Indices for the Branching of Tree-like Graphs.
146. C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. University of Illinois Press, Urbana, IL, 1949.
147. L. Brillouin, *Science and Information Theory*. Academic, New York, 1962.
148. D. Bonchev, *Information Theoretic Indices for Characterization of Chemical Structure*. Wiley, Chichester, England, 1983.
149. L. B. Kier, *J. Pharmacol. Sci.*, **69**, 807 (1980). Use of Molecular Negentropy to Encode Structure Governing Biological Activity.
150. T. Y. Shen, in *Top. Med. Chem.*, Vol. 1, J. L. Rabinowitz and R. M. Myerson, Eds., Interscience Publishers, New York, 1967, p. 19. Anti-Inflammatory Agents.
151. P. L. Luisi, *Nature (London)*, **64**, 569 (1977). Molecular Conformational Rigidity: An Approach to Parameterization.
152. K. S. Pitzer, *J. Am. Chem. Soc.*, **77**, 3427 (1955). The Volumetric and Thermodynamic Properties of Fluids. I. Theoretical Basis and Virial Coefficients.
153. C. A. Passut and R. P. Danner, *Ind. Eng. Chem. Process Des. Dev.*, **12**, 365 (1973). Acentric Factor. A Valuable Calculating Parameter for the Properties of Hydrocarbons.

154. L. B. Kier, *Quant. Struct.-Act. Relat.*, **6**, 8 (1987). Inclusion of Symmetry as a Shape Attribute in Kappa Index Analysis.
155. L. B. Kier, *Quant. Struct.-Act. Relat.*, **6**, 117 (1987). A Substituent Steric Effect Based on the Molecular Graph.
156. K. J. Shah and E. A. Coats, *J. Med. Chem.*, **20**, 1001 (1977). Design, Synthesis, and Correlation Analysis of 7-Substituted 4-Hydroxyquinidine-3-carboxylic acids.
157. L. H. Hall, E. L. Maynard, and L. B. Kier, *Environ. Toxicol. Chem.*, **8**, 783 (1989). QSAR Investigations of Benzene Toxicity to Fathead Minnow Using Molecular Connectivity.
158. L. H. Hall and L. B. Kier, *Tetrahedron*, **33**, 1953 (1977). A Molecular Connectivity Study of Electron Density in Alkanes.
159. MOLCONN-X is a FORTRAN program, a new version of MOLCONN2, that computes molecular connectivity chi indexes, kappa shape indexes, subgraph counts, and other quantities of topological interest. Input is by one of several forms of connection table and output is available in several forms of files. For information, contact L. H. Hall.
160. M. Randić, G. M. Brissey, and C. L. Wilkins, *J. Chem. Inf. Comput. Sci.*, **21**, 52 (1981). Computer Perception of Topological Symmetry via Canonical Numbering of Atoms.
161. M. Randić, D. O. Oakland, and D. J. Klein, *Comput. Chem.*, **7**, 35 (1986). Symmetry Properties of Chemical Graphs. IX. Valence Tautomerism in the Heptaphosphorous Skeleton.
162. R. E. Carhart, *J. Chem. Inf. Comput. Sci.*, **17**, 113 (1977). Canonical Numbering and Constitutional Symmetry.
163. F. M. Berger, C. V. Hubbard, and B. J. Ludwig, *Appl. Microbiol.*, **1**, 146 (1953). The Antimicrobial Action of Certain Glycerol Ethers and Related Compounds.
164. C. Hansch and E. J. Lien, *J. Med. Chem.*, **14**, 653 (1971). Structure-Activity Relationships in Antifungal Agents. A Survey.