

Intermolecular Accessibility: The Meaning of Molecular Connectivity

Lemont B. Kier*

Department of Medicinal Chemistry, Virginia Commonwealth University, Richmond, Virginia 23298

Lowell H. Hall

Department of Chemistry, Eastern Nazarene College, Quincy, Massachusetts 02170

Received November 11, 1999

The molecular connectivity indices are shown to be the numerical possibilities of a molecule encountering another identical molecule, derived from the bond accessibilities of each. The meaning of the δ values and the significance of all suppressed X–H bonds is described. A new concept of the meaning of molecular connectivity, built around bimolecular encounter accessibility, is presented.

INTRODUCTION

A quarter of a century has passed since Milan Randić proposed an algorithm to encode bond contributions to a molecular branching index.¹ For the first time it became possible to offer quantitative statements about the extent of branching in a molecule. Beyond that, Randić demonstrated that alkanes could be ordered by this scheme to correlate with a physical property, the boiling point. This branching algorithm formed the basis of a structure description paradigm called molecular connectivity developed over the next decade by Kier and Hall.^{2,3} This is now a very widely used nonempirical structure description in quantitative structure–activity analyses and model development.⁴

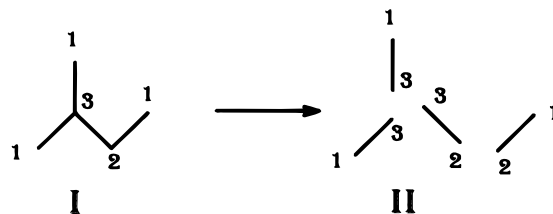
A salient feature of molecular connectivity is that it encodes structure in a nonempirical way. It is not a measured property, nor does it derive from or directly translate into a particular physical property. This has, strangely, not been understood by some investigators employing structure–activity models. In this paper, we will examine the Randić algorithm, propose a concept of the significance of the bond codes, and offer an interpretation of the information encoded by molecular connectivity.

THE RANDIC ALGORITHM

The molecule is considered to be a sum of parts, namely, the bonds connecting pairs of atoms. Each atom in a molecule is encoded by a cardinal number, δ , the count of all bonded atoms other than hydrogen. In chemical graph theory terms, the molecule is a hydrogen-suppressed graph illustrated by isopentane, I. In common usage the hydrogen atoms are collected with the attached atom as a hydride group such as CH₃ and NH₂ and are represented by a single δ value. The molecule is dissected into fragments or bonds, each retaining the δ values assigned in the original graph as shown for isopentane, II. This decomposition produces a set of fragments encoded by the two δ values of the atoms comprising each bond, shown in Table 1.

Table 1. Bond Types in Isopentane

bond type (δ_i, δ_j)	bond type code	$\delta_i \delta_j$	C_{ij}
1, 2	a	2	0.707
1, 3	b	3	0.577
2, 3	c	6	0.408



Randić proceeded to reduce these bond descriptors into a single number, encoding the information about the differences between the fragments. He elected to use the product of the δ values, $(\delta_i \delta_j)$, taken to the -0.5 power. Stepwise, this is the product, the square root, and the reciprocal. These values are shown in Table 1. Kier and Hall^{5,6} have designated this as the C_{ij} value:

$$C_{ij} = (\delta_i \delta_j)^{-0.5}$$

The final step in the creation of a molecular index is to recombine these values by summation. This index was proposed by Randić as a branching index and was shown to correlate well with alkane boiling points.¹ Kier and Hall² later generalized this as $^1\chi$, a chi index of the first order:

$$^1\chi = \sum (\delta_i \delta_j)^{-0.5} \quad \text{summed over all skeletal bonds}$$

Subsequent developments of this paradigm have been covered elsewhere.^{2,3}

Any attempt at interpreting the structure information inherent in the χ indices must begin with a close examination of the indices characterizing the bonds. Expanding the list in Table 1 to include all bonds possible in alkanes, we have the 10 entries in Table 2. The first bond type entry, (1, 1), is unique for ethane among the alkanes. The others are found

Table 2. Bond Types in C_{sp^3} Molecules

bond type (δ_i, δ_j)	$\delta_i \delta_j$	$C_{ij} = A_{ij}$
1, 1	1	1.000
1, 2	2	0.707
1, 3	3	0.577
1, 4	4	0.500
2, 2	4	0.500
2, 3	6	0.408
2, 4	8	0.354
3, 3	9	0.333
3, 4	12	0.289
4, 4	16	0.250

in the whole set of alkanes and in the skeletons of all covalently bonded molecules. The bonds shown in Table 2 are listed according to the decreasing value of C_{ij} that corresponds to an increase in branching at one or both of the atoms comprising the bond. The C_{ij} index is thus successful in encoding this attribute.

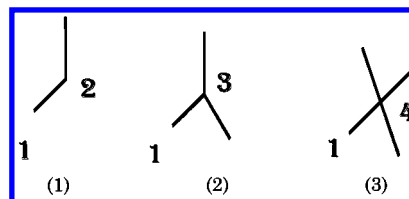
We now address the question of the significance of this algorithm and the structural interpretation of C_{ij} . Clearly C_{ij} parallels the degree of branching associated with each fragment. The algorithm to calculate this value was not derived; it was stated. The objective was to encode the influence of the branched state of each atom in the bond on the relative contribution of that bond to the whole molecule. In other words, the characterization of some aspect of the structure of a molecule was the goal in assigning a numerical value that might correlate with some measured properties.

A BIMOLECULAR ENCOUNTER MODEL

An alternate view of the problem of defining molecular structure and relating this to the value of a physical property is considered. This view focuses on the bond indices C_{ij} , while conventional wisdom has been directed at the numerical descriptors of whole molecules. The basic premise of this alternative is that a useful description of molecular structure represents the molecule in a milieu of other molecules. This idea was expressed by Seybold,^{7,8} who wrote about the structure influence on packing of molecules in their milieu, an attribute dependent upon their shape. Our view leads to a quality of intermolecular dispersion that would account for several physical properties. In fact, a measurement of a physical property reflects the collective influence of encounters of each molecule with other molecules in its immediate environment. Our model places emphasis on the encounters among molecules, in particular the possibilities of fragments of one encountering fragments of another. Kier and Hall have focused on this concept by exploring the possibilities of encounters among two molecules.⁹ This bimolecular encounter model reflects in a microcosm what is very likely taking place in a manifold of molecules in a bulk system. Analysis of these encounters form the basis of an interpretation of parts of the algorithm introduced by Randic.

ACCESSIBILITY INTERPRETATION OF BOND FRAGMENTS

The δ Values. We propose that the significance of a C_{ij} value is its encoding of the relative accessibility of that bond to encounter another bond in an intermolecular interaction. The term accessibility is defined as the topological and

**Figure 1.** Molecular fragments: (1) δ values of 1 and 2, (2) δ values of 1 and 3, and (3) δ values of 1 and 4.

electronic availability of one bond to engage in some interaction with a bond in another molecule. How then can we develop *de novo*, numerical expressions for accessibility in this context? Consider three hydrogen-suppressed graph fragments of molecules with a common feature, a methyl group, Figure 1. Each atom (hydride group) in each fragment is designated by a δ value, which is the count of σ electrons contributed to adjacent bonds other than hydrogen. The atoms in Figure 1 with $\delta = 1$ contribute all of their non-hydrogen σ electrons to the bond. In fragment 1, one atom has a $\delta = 2$ value, denoting the contribution of $1/2$ of its non-hydrogen σ electrons to the bond with the methyl group. By analogy, fragment 2 has an atom designated $\delta = 3$, signifying that only $1/3$ of its non-hydrogen σ electrons are contributed to the bond shown. For fragment 3 it follows that $1/4$ of the σ electrons are contributed to the bond from the atom with $\delta = 4$. The δ values therefore have a dual meaning. First the δ value is the count of neighboring atoms bonded to an atom in the hydrogen-suppressed graph, corresponding to the count of the σ electrons contributed by that atom to bonded, non-hydrogen atoms. Second, the reciprocal of the δ value, $1/\delta$, is the fraction of the total number of non-hydrogen σ electrons contributed to each bond formed with that particular atom.

The Bond Terms. The $1/\delta$ terms described above may be viewed as effective contributions to the relative accessibility of the bond in which they are a part. Galvez¹⁰ has compared these δ values to molecular orbital parameters and has derived a relationship between the bond order and the $(1/\delta_i 1/\delta_j)^{1/2}$ value. From this we interpret the term to be a relative bond accessibility value that we label A_{ij} . All possible alkane bond A_{ij} values are shown in Table 2. When this value is high for a bond, there is an expectation that the bond is relatively accessible to other bonds in the milieu. Conversely, a low value of A_{ij} infers a low accessibility. Looking at Figure 1 again, we conclude that the bond in fragment 1 has a greater accessibility to other bonds in its environment than does fragment 2 or 3.

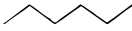
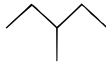
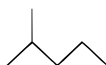
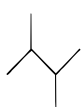

BIMOLECULAR ENCOUNTER PARAMETERS

Most physical properties arise from intermolecular encounters of molecules, governed by the pattern of atoms and bonds, primarily on the mantle of each molecule. Kier and Hall have used the counts of the numbers of bimolecular encounters and have formulated a series of parameters encoding a relationship with some physical properties.⁹ Expanding on that work, we propose a model based on the encounters of bonds in two molecules using isopentane, I, as an example. The encounter of two molecules of isopentane is considered to be some function of the interaction of the bonds between molecules. A series of bimolecular bond encounter parameters is calculated from the manifold of bond

Table 3. Intermolecular Encounter Possibilities

	1, 2	1, 3	2, 2 1, 4	2, 3	2, 4	3, 3	3, 4	4, 4
1, 2	0.50	0.41	0.35	0.29	0.25	0.24	0.20	0.18
1, 3		0.33	0.29	0.24	0.20	0.19	0.17	0.14
2, 2/1, 4			0.25	0.20	0.18	0.17	0.15	0.13
2, 3				0.17	0.14	0.13	0.12	0.10
2, 4					0.13	0.12	0.10	0.09
3, 3						0.11	0.10	0.08
3, 4							0.08	0.07
4, 4								0.06

Table 4. Bimolecular Encounter Accessibilities for Hexane Isomers

Hexane Isomer	$\sum_m(A_{ij})_m \sum_n(A_{kl})_n$	${}^1\chi = P_m$
	8.491	2.914
	7.885	2.808
	7.328	2.707
	6.985	2.643
	6.559	2.561

accessibilities operating between two molecules of isopentane. In this model, the encounter accessibility of any two bonds on two different molecules is calculated as the product of the individual accessibilities, A_{ij} , of each. Table 3 shows all possible bimolecular bond encounter terms for any saturated hydrocarbon molecule.

To describe this in another way, the encounter of two isopentane molecules includes the possible interaction of the two fragments of the type (1, 2) shown in Figure 1 (1). The accessibility, A_{ij} , of each of these bonds is 0.707 from Table 2. The encounter possibility of two such bonds on separate molecules is assumed to be the product of the accessibilities, $(0.707)(0.707)$, or 0.5. The complete set of possible encounter accessibilities of two isopentane molecules can be extracted from Table 3. The products of all such interactions, for two isopentane molecules, $A_{ij}A_{kl}$, from Table 3, is summed to give a total bimolecular encounter possibility for isopentane. From the data in Table 3 we calculate the total bimolecular interaction possibility, P_T , for two isopentane molecules, m and n , to be:

$$\sum_m(A_{ij})_m \sum_n(A_{kl})_n = 5.153$$

An expanded list of P_T values for all hexane isomers is shown in Table 4.

The bimolecular encounter possibility, P_T , is the product of the attributes of each molecule, P_m and P_n , in the

encounter. In the case of isopentane, the contribution to P_T from molecule m (and also molecule n) is

$$P_m = 2.270.$$

COMPARISON OF THE TWO MODELS

The Randic algorithm computes a term for each bond type and then sums these to give a molecular connectivity index. Using isopentane as our example, we assign to each different bond type a letter code, designating a different bond type, shown in Table 1. The three different bond types for isopentane are coded a, b, and c. The appropriate summation of these indices in isopentane gives the molecular connectivity index of the first order. The calculation for isopentane is

$${}^1\chi = 2b + a + c \quad (1)$$

We now consider our bimolecular model in which we compute each possible bond encounter as a product of bond accessibilities among two molecules of isopentane. Using the same codes for the bond types in Table 1, all possible encounters of bonds among two isopentane molecule are

$$bb, bb, bb, bb = 4bb = 4b^2$$

$$ba, ba, ba, ba = 4ba$$

$$bc, bc, bc, bc = 4bc$$

$$ac, ac = 2ac$$

$$aa = a^2$$

$$cc = c^2$$

Summing these encounters, we arrive at the total bimolecular interaction possibility for two isopentane molecules, m and n :

$$P_T = P_m P_n = \sum_m(A_{ij})_m \sum_n(A_{kl})_n = 4b^2 + 4ba + 4bc + 2ac + a^2 + c^2 \quad (2)$$

It follows that the interaction possibility of one isopentane molecule P_m is

$$P_m = 2b + a + c$$

To confirm this, the sum of bimolecular interaction possibilities for isopentane is 5.153, and the contribution of one molecule, m , is $P_m = 2.270$.

We conclude that the molecular connectivity index is the contribution of one molecule to the bimolecular interactions arising from encounters of bonds among two molecules. The molecular connectivity index ${}^1\chi = P_m$.

This result is confirmed by relating the ${}^1\chi$ indices of other molecules to the P_m values from their bimolecular interaction values as shown in Table 4. It can be shown that the extended molecular connectivity indices^{2,3,11} are equal to the $P_{i..n}$ values of bimolecular interactions of higher order fragment types.

In the simplest case, the graph of a molecule with only one bond (type (1, 1)) has an accessibility value of x , while the bimolecular accessibility value is $\sum_m(A_{ij})_m \sum_n(A_{kl})_n = x^2$. The equality, $x = x^2$, can only exist for positive values when

$x = 1$. This is the only ${}^1\chi$ value allowed for a molecular graph of one bond, in the simple case, the graph for ethane.

VALENCE δ INTERPRETATION

The valence delta value, δ^V , introduced by Kier and Hall¹² is defined as

$$\delta^V = Z^V - H$$

where Z^V is the count of valence electrons in the atom. The A_{ij} value for a C–O bond in ethanol is thus derived from the two δ^V values; $\delta^V = 2$ for the C and $\delta^V = 5$ for the OH group. The A_{ij} value is therefore

$$A_{ij} = [(2)(5)]^{-0.5} = 0.316$$

This is a lower value than the A_{ij} for a comparable C–C bond, where $A_{ij} = 0.707$.

The lower A value reflects the polarity of the C–O bond due to the unequal electronegativity. In the χ formalism the C–O bond is estimated to have a lower accessibility for interactions with other bonds in other molecules. The important bimolecular interacting potential of the C–O bond comes from the much greater accessibility of the electron lone pairs on the OH fragment. Pogliani¹³ has demonstrated that the δ^V values correlate with the atomic charge distributions, and so we expect a much greater interaction potential from those atoms with π and lone pair electrons. This electron accessibility has been shown to be encoded in the electrotopological state index (E-state).⁹

THE C–H BONDS

A question now arises about the role of the C–H bonds, suppressed in the creation of the chemical graph structure for the molecular connectivity calculations. Since the C–H bonds are the prominent mantle features of the alkanes we have considered, they must be an important part of any bimolecular or intermolecular encounters. Accordingly, their role must be addressed in any consideration of the meaning of molecular connectivity indices. The following argument links them directly to the accessibility model shown above.

The count of C–H bonds in a molecule is equal to the count of H atoms. More generally, X–H bonds can be counted by counting the H atoms. The definition of a δ value for a carbon atom in a molecular graph is

$$\delta = 4 - H$$

where H is the count of hydrogen atoms. It follows that

$$1/\delta = (4 - H)^{-1}$$

The A_{ij} of a bond, hence the accessibility, can be written in terms of the count of H ; thus

$$A_{ij} = [(4 - H)_i(4 - H)_j]^{-0.5}$$

From this we calculate the bimolecular encounter possibilities as

$$\Sigma\{[(4 - H)_i(4 - H)_j]^{-0.5}[(4 - H)_i(4 - H)_j]^{-0.5}\}$$

This is a result equivalent to eq 2 using δ values. This establishes a direct relationship between the count of interacting C–H bonds and the molecular connectivity index for the molecule.

CONCLUSION

The results of this study lead to an interpretation directly in terms of molecular structure for the molecular connectivity indices. After a quarter of a century of productive use, we can support our claim that the molecular connectivity indices are indeed nonempirical structure descriptions that are rich in information. This study also allows us to lay claim to finding a fundamental meaning of the χ indices, which is the encoding of intermolecular accessibility. It is now clear why these indices are so useful in quantifying the relationship between structure and physical properties, particularly those that arise from intermolecular interactions. This realization should, finally, correctly categorize these indices in their rightful place in the pantheon of quantitative descriptions of molecular structure.

REFERENCES AND NOTES

- (1) Randic, M. On Characterization of Molecular Branching. *J. Am. Chem. Soc.* **1975**, 97, 6609–6615.
- (2) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Chemistry and Drug Research*; Academic Press: New York, 1976.
- (3) Kier, L. B.; Hall, L. H. *Molecular Connectivity in Structure Activity Analysis*; Research Studies Press: Letchworth, England, 1986.
- (4) *Molconn-Z* (version 3.15); Hall Associates Consulting, 2 Davis St., Quincy, MA 02170.
- (5) Hall, L. H.; Kier, L. B.; Murray, W. J. Molecular Connectivity II: Relationship to Water Solubility and Boiling Point. *J. Pharm. Sci.* **1975**, 64, 1974–1979.
- (6) Kier, L. B.; Murray, W. J.; Hall, L. H. Molecular Connectivity. 4. Relationships to Biological Activities. *J. Med. Chem.* **1975**, 18, 1272–1274.
- (7) Seybold, P. G.; May, M.; Bagal, U. A. Molecular Structure–Property Relationships. *J. Chem. Educ.* **1987**, 64, 575–581.
- (8) Needham, D. E.; Wei, I.-Chien; Seybold, P. G. Molecular Modelling of the Physical Properties of the Alkanes. *J. Am. Chem. Soc.* **1988**, 110, 4186–4194.
- (9) Kier, L. B.; Hall, L. H. *Molecular Structure Description: The Electrotopological State*; Academic Press: New York, 1999; pp 175–178.
- (10) Galvez, J. On the Topological Interpretation of Electronic and Vibrational Molecular energies, *J. Mol. Struct.: THEOCHEM* **1998**, 429, 255–264.
- (11) Kier, L. B.; Murray, W. J.; Randic, M.; Hall, L. H. Molecular Connectivity V: Connectivity Series Concept Applied to Density. *J. Pharm. Sci.* **1976**, 65, 1226–1229.
- (12) Kier, L. B.; Hall, L. H. Molecular Connectivity VII: Specific Treatment of Heteroatoms. *J. Pharm. Sci.* **1976**, 65, 1806–1820.
- (13) Pogliani, L. The Molecular Connectivity Method: A Powerful Tool in the Study of Biologically Relevant Molecules. *Curr. Top. Peptide Prot. Res.* **1994**, 1, 119–134.

CI990135S