CROATICA CHEMICA ACTA CCACAA **75** (2) 371–382 (2002)

ISSN-0011-1643
CCA-2806 *Original Scientific Paper*

The Meaning of Molecular Connectivity: A Bimolecular Accessibility Model*

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

[a] *Center for the Study of Biological Complexity, Virginia Commonwealth University, Richmond, Virginia 23298, USA*

[b] *Department of Chemistry, Eastern Nazarene College, Quincy, Massachusetts 02170, USA*

Received March 19, 2001; revised May 4, 2001; accepted June 5, 2001

The simple molecular connectivity indices are analysed for the information in the bond terms. It is found that these terms reflect the relative accessibility of each bond to encounter other bonds of the same molecule in a milieu. The total possibility of each molecule to encounter another molecule in a bimolecular interaction is found to be the molecular connectivity index for that molecule. The molecular connectivity indices are interpreted to be the bimolecular interaction possibilities of a molecule in a milieu.

Key words: molecular connectivity, intermolecular accessibility, bimolecular encounters, meaning of molecular connectivity.

MOLECULAR STRUCTURE

The development of new drugs with pharmacological efficacy and clinical utility is now a major activity in the chemical and biological sciences. Successes in mid-century have led to a level of confidence in our ability to make some predictions in the design of molecules with desirable characteristics. Today we are on the threshold of rational drug design based on the occasional ability to recreate and model the molecular-level scene of action of a ligand molecule and an effector using computer graphic simulations. From

---

* Offered in recognition of the creativity of Milan Randić and his 70th year.

** Author to whom correspondence should be addressed. (E-mail: kier@hsc.vcu.edu)

such models, variations in structure may be made with the objective of improving the drug-receptor encounters leading to a better drug.

The opportunity to use this approach in drug design is limited to situations in which there is considerable information such as the specific sites of action or the structure of the enzymes involved and the ligands that interact. In the absence of a well-defined molecular target, it is necessary to adopt other strategies to develop a drug. Beyond the attribute of biological activity, in the realm of pharmaceutical properties such as absorption, distribution, metabolism and excretion, it is not possible to build a model around a macromolecular system with an active site. These properties are in the domain of molecular systems where something resembling an effector is an evanescent group of water molecules in intimate contact with solutes or protein surface fragments forming a complex system.[1]

This reflection leads us to the consideration of an alternative approach to drug design, one that we have called *quantitative information analysis*.[2] In this approach a series of molecules is presented to a biological system and the properties of interest are measured. We are concerned with the structure of the ligand, A, and the numerical readout from system B. The intervening systems are a hierarchy of complex systems, each with emergent properties fueling the creation of still higher orders of complex systems. Ultimately a system yields the emergent properties that we measure, B, and that are clinically useful. The essence of this approach to drug design is the creation of a model relating some attribute of the molecules introduced to the biological system and the measured properties emerging from the encounter. We make the general statement that $B = f(A)$. We do not expect to acquire much wisdom in the form of mechanistic models of many of the intervening events. This is the realm of molecular biology. Our objective is to modify the molecular structures to improve the measured output in a conservative model.

The process of quantitative information analysis has followed two paths over the years. One path has lead to the use of physical properties to describe a molecule in a model, relating it to a measured response.[3] This has come to be called quantitative structure-activity relationship (QSAR), where the word »structure« was loosely used to denote measured or estimated physical properties. At the same time, an alternative paradigm to physical property models arose in the form of theoretical models of structure derived from molecular orbital theory.[4] These models attempted to relate a structure quantitation to measured responses in an effort to create a predictive model. Since that time, innovations to the concept of structure have emerged that extend the practical utility of QSAR models. In particular the introduction of a practical graph-based topological index by Randić[5] a quarter of a cen-

tury ago and developed by Kier and Hall,[6,7] made possible a description of structure that is simple and demonstratively valuable in predictive power.

It is important to distinguish between physical properties and structure concepts that have often become blurred. Attempts to rectify this misunderstanding were made by Norrington[8] who proposed the term *property-activity relationships* as a distinct sub-class of structure-activity models. Hoffmann[9] has written eloquently on this subject from a broad chemical perspective.

We regard molecular structure as a collective term for codes by which we describe a molecule in quantitative terms. It is a model of the form of a molecule, which produces a series of functions called properties. The functions or properties are measured attributes, which, in the case of molecules, are averaged values of responses to chemical or physical input into a system containing the molecule in question. The numerical values of the properties present a mosaic of information about the system. From this information we weave a model of what the structure may be that gave birth to our measurements. Structure is a model; it is a presumption of what is *there*, functioning in response to our physical inquiry. A representation of what is there may take the form of statements of content, probability, accessibility, topology, relationships, or complexity. In this sense, any representation may be incomplete yet judged adequate for a particular purpose. The function or properties of a molecule are dependent upon the form or structure. It is an immutable relationship; structure is the antecedent to properties; form precedes functions.

MOLECULAR CONNECTIVITY

The Randić Branching Index

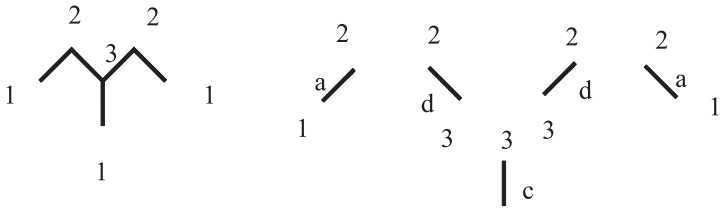
Twenty five years have passed since Milan Randić proposed an algorithm to encode bond contributions to a molecular branching index.[5] From this effort it has become possible to offer quantitative statements about the extent of branching in a molecule. Beyond that, Randić demonstrated that alkanes could 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.[6,7] This is now a very widely used non-empirical structure description in quantitative structure-activity analyses and model development.[10]

The essence of molecular connectivity is the encoding of structure in a non-empirical 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 an alternate structure concept, and offer an interpretation of the information encoded by molecular connectivity.

The Randić Algorithm

The molecule is regarded as a sum of 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 3-methylpentane, **I**. In common notation the hydrogen atoms are collected with the attached atom and represented as a hydride group such as CH_3 and CH_2 and are designated by an appropriate δ value. The molecule is dissected into frag-



ments or bonds, each retaining the δ values assigned in the original graph as shown for 3-methylpentane, **II**. This decomposition produces a set of fragments encoded by the two δ values of the atoms comprising each bond, shown in Table I.

Randić combined 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 I. Kier and Hall[5,6] have designated this as the C_{ij} value:

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

The step leading to the creation of a molecular index was the combination of these values by summation. This index was proposed by Randić as a

branching index and was shown to correlate well with alkane boiling points.[5] Kier and Hall[6] later generalized this as, $^1\chi$, a chi index of the first order:

$$^1\chi = \Sigma (\delta_i \delta_j)^{-0.5}$$

summed over all skeletal bonds.

Interpreting the structure information inherent in the chi indices begins with an examination of the indices characterizing the bonds. All bond types possible in alkanes are the ten entries in Table I. The first bond type, (1,1), is unique for ethane among the alkanes. The others are found among all of the alkanes, and in the non-atom specific skeletons of covalently bonded molecules. The bonds shown in Table I are listed according to the decreasing value of C_{ij} which corresponds to the perception of an increase in branching at one or both of the atoms comprising the bond. The C_{ij} index is thus successful in ranking this attribute.

TABLE I

Bond types in Csp^3 molecules

Bond Type (δ_i, δ_j)	Code	$\delta_i\ \delta_j$	$C_{ij} = A_{ij}$
1,1	unique	1	1.000
1,2	a	2	0.707
1,3	c	3	0.577
1,4	f	4	0.500
2,2	b	4	0.500
2,3	d	6	0.408
2,4	g	8	0.354
3,3	e	9	0.333
3,4	h	12	0.289
4,4	i	16	0.250

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 which might correlate with some measured properties.

A BIMOLECULAR ENCOUNTER MODEL

Accessibility Interpretation of Bond Fragments

Another approach to the definition of molecular structure and its relationship to the value of a physical property is now proposed. This approach focuses on the bond indices C_{ij}, while the 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[11] who wrote about the structure influence on packing of molecules in their milieu, an attribute dependent upon their shape. This 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 their 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.[12] 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 forms the basis of our interpretation of the molecular connectivity index.

The Delta Values

Our interpretation of the significance of a C_{ij} value is that it encodes the relative accessibility of a bond to encounter another bond in another molecule, leading to an intermolecular interaction. The term accessibility is defined as the topological and 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, in structure **II**. Each atom (hydride group) in each fragment is designated by a delta value, which is the count of sigma electrons contributed to bonds formed with adjacent atoms. The atoms in **II** with $\delta = 1$ contribute all of their non-hydrogen sigma electrons to the bond. In the fragment **IIa**, one atom has a $\delta = 2$ value, denoting the contribution of 1/2 of its non-hydrogen sigma electrons to the bond with the methyl group. By analogy, the fragment **IIc** has an atom designated $\delta = 3$, signifying that only 1/3 of its non-hydrogen sigma electrons are contributed to the bond shown. For any fragment with $\delta = 4$, it follows that 1/4 of the sigma electrons are contributed to each of the bonds of which it is a part.

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. This corresponds to the count of the sigma 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 sigma electrons contributed to each bond formed with that particular atom.

The Bond Terms

The $1/\delta$ terms reflect effective contributions of atoms to the relative accessibility of the bond of which they are a part. Galvez [13] has compared these delta 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 $(1/\delta_i 1/\delta_j)^{1/2}$ term to be a relative bond accessibility value which we relabel A_{ij}. All possible alkane bond A_{ij} values are shown in Table I. 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. From structure **II**, we conclude that the bond in fragment **IIa** has a greater accessibility to other bonds in its environment than do fragments **IIc** or **IId**.

Bimolecular Encounter Parameters

Intermolecular encounters of molecules, governed by the pattern of atoms and bonds, influence many physical property values. Kier and Hall have studied the counts of the number of possible bimolecular bond encounters and have formulated a series of parameters encoding a relationship with some physical properties.[12] Expanding on that study we propose a model based on the encounters of bonds in two molecules using 3-methylpentane, **I**, as an example. The encounter of two molecules of 3-methylpentane 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 accessibilities operating between two molecules of 3-methylpentane. 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 II shows a matrix of all possible bimolecular bond encounter terms for any alkane.

Stated another way, the encounter of two 3-methylpentane molecules includes the possible interaction of the two fragments of the type (1,2) shown in structure **IIa**. The accessibility, A_{ij}, of each of these bonds is 0.707 from Table II. The possibility of the encounter of two such bonds on separate molecules is evaluated as the product of the accessibilities, (0.707)(0.707) or 0.5.

TABLE II

Intermolecular encounter possibilities

Bond	Types	a 1,2	c 1,3	b/f 2,2/1,4	d 2,3	g 2,4	e 3,3	h 3,4	i 4,4
a	1,2	0.50	0.41	0.35	0.29	0.25	0.24	0.20	0.18
c	1,3		0.33	0.29	0.24	0.20	0.19	0.17	0.14
b/f	2,2/1,4			0.25	0.20	0.18	0.17	0.15	0.13
d	2,3				0.17	0.14	0.13	0.12	0.10
g	2,4					0.13	0.12	0.10	0.09
e	3,3						0.11	0.10	0.08
h	3,4							0.08	0.07
i	4,4								0.06

The complete set of encounter accessibilities of two 3-methylpentane molecules can be extracted from Table II. The products of all such interactions, $(A_{ij})(A_{kl})$, for two 3-methylpentane molecules are summed to give a total bimolecular encounter possibility for 3-methylpentane. From the data in Table II we calculate the total bimolecular interaction possibility, P_T, for two 3-methylpentane molecules, m and n, to be:

$$\Sigma_m(A_{ij})_m\Sigma_n(A_{kl})_n = 7.885 \ .$$

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 3-methylpentane the contribution to P_T from molecule, m, (and also molecule n) is P_m = 2.808.

COMPARISON OF THE TWO MODELS

The Equivalence of $^1\chi$ *and* P_m

The algorithm of Randić computes a term for each bond type and then sums these to give a molecular connectivity index. Using 3-methylpentane as our example, we assign to each different bond type a letter code, designating a different bond type, shown in Table I. The three different bond types for 3-methylpentane are coded a, c, and d. The appropriate summation of these indices in 3-methylpentane gives the molecular connectivity index of the first order, $^1\chi$. The calculation for 3-methylpentane is:

$$^1\chi = 2a + c + 2d \ . \qquad (1)$$

In our bimolecular interaction model we compute each possible bond encounter as a product of bond accessibilities among two molecules of 3-methylpentane. Using the same codes for the bond types in Table I, all possible encounters of bonds among two 3-methylpentane molecules are:

$$\begin{aligned} \text{aa, aa, aa, aa} &= 4a^2 \\ \text{dd, dd, dd, dd} &= 4d^2 \\ \text{ad, ad, ad, ad, ad, ad, ad, ad} &= 8ad \\ \text{ac, ac, ac, ac} &= 4ac \\ \text{cd, cd, cd, cd} &= 4ac \\ \text{cc} &= c^2 \end{aligned}$$

The sum of the number of bond encounters is the product of the number of bonds in each. The interaction possibility of one 3-methylpentane molecule P_m is:

$$P_m = 2a + c + 2d \ .$$

The total interaction possibility, P_T, is the product of the two interaction possibilities for molecules m and n:

$$P_T = P_m P_n = \Sigma(A_{ij})(A_{kl}) = 4a^2 + 4ac + 8ad + 4ac + 4d^2 + c^2 \ . \qquad (2)$$

To test this, the sum of bimolecular interaction possibilities for 3-methylpentane is 7.885, and the contribution of each molecule, m or n, is P_m = 2.808. The equality exists:

$$7.885 = (2.808)(2.808) \ .$$

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

This result is confirmed by relating the ${}^1\chi$ indices of other molecules, y, to the P_y values from their bimolecular interaction values as shown in Table III. It can be shown that the extended molecular connectivity indices[6,7] are equal to the $P_{i...n}$ values of bimolecular interactions of higher order fragments.

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, $\Sigma_m(A_{ij})_m\Sigma_n(A_{kl})_n = x^2$. The equality, $x = x^2$ can only exist for positive values when $x = 1$. This is the only ${}^1\chi$ value possible for a molecular graph of one bond, in the simple case, the graph for ethane.

TABLE III

Bimolecular encounter accessibilities for hexane isomers

Hexane isomer	$\Sigma(A_{ij})(A_{kl})$	${}^1\chi = P_{ij}$
	8.491	2.914
	7.885	2.808
	7.328	2.707
	6.985	2.643
	6.559	2.561

The C-H Bonds

The role of the C-H bonds, suppressed in the creation of the chemical graph structure for the molecular connectivity calculations, can now be considered. 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 encounter. Accordingly, their role must be addressed in any consideration of the meaning of molecular connectivity indices for the alkanes. 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: $\delta = 4 - \mathrm{H}$ where H is the count of hydrogen atom. It follows that:

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

The accessibility, A_{ij}, of a C-H bond can be written in terms of the count of H atoms, thus:

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

From this we calculate the bimolecular encounter possibilities as:

$$\sum\{[(4\text{-H})_i(4\text{-H})_j]^{-0.5}\ [(4\text{-H})_i(4\text{-H})_j]^{-0.5}\} \ .$$

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.

DISCUSSION

The results of this study lead to a detailed structural interpretation of the molecular connectivity indices. After a quarter of a century of productive use, we can support our concept that molecular connectivity indices are indeed non-empirical structure descriptions that are rich in information. This study allows us to lay claim to finding a fundamental meaning of the chi indices, that is the encoding of bimolecular encounter accessibility. It is now clear why these indices are useful in quantifying the relationship between structure and physical or biological 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 descriptors of molecular structure.

REFERENCES

1. L. B. Kier and B. Testa, *Adv. Drug Res*. **26** (1995) 1–43.
2. L. B. Kier and L. H. Hall, *Med. Chem. Res*. **7** (1997) 335–339.
3. C. Hansch, *Acc. Chem. Res.* **2** (1969) 232–239.
4. L. B. Kier, *Molecular Orbital Theory in Chemistry and Drug Research,* Academic Press, New York, 1971.
5. M. Randić, *J. Am. Chem. Soc.* **97** (1975) 6606–6615.
6. L. B. Kier and L. H. Hall, *Molecular Connectivity in Chemistry and Drug Research*, Academic Press, New York, 1976.
7. L. B. Kier and L. H. Hall, *Molecular Connectivity in Structure Activity Analysis*, John Wiley & Sons, New York, 1986.
8. F. E. Norrington, R. M. Hyde, S. G. Williams, and R. Wooten, *J. Med. Chem.* **18** (1975) 604–608.
9. R. Hoffmann and P. Laszlo, *Angew. Chem.* **30** (1991) 1–16.
10. L. H. Hall, Molconn-Z (version 3.15), Hall Associates Consulting, 2 Davis St., Quincy, MA 02170.
11. P. G. Seybold, M. May, and U. A. Bagal, *J. Chem. Ed.* **64** (1987) 575–581.
12. L. B. Kier and L. H. Hall, *Molecular Structure Description: The Electrotopological State*, Academic Press, San Diego, CA, 1999.
13. J. Galvez, *J. Molec. Struct.* **429** (1998) 255–264.

SAŽETAK

Značenje molekulske povezanosti: model bimolekulske dostupnosti

Lemont B. Kier i Lowell H. Hall

Jednostavni indeksi molekulske povezanosti analizirani su s pomoću doprinosa svake veze u molekuli. Nađeno je da ti članovi reflektiraju relativnu pristupačnost svake veze pri dodiru s drugim vezama iste molekule. Nađeno je da je ukupna vjerojatnost sudara jedne molekule s drugom u bimolekulskoj reakciji jednaka indeksu molekulske povezanosti za tu molekulu. Indeksi molekulske povezanosti interpretirani su kao vjerojatnosti bimolekulske interakcije molekule u reakcijskoj sredini.