AN INDEX OF ELECTROTOPOLOGICAL STATE FOR ATOMS IN MOLECULES

Lemont B. KIER
Department of Medicinal Chemistry, Virginia Commonwealth University, Richmond, VA 23298, USA
Lowell H. HALL
Department of Chemistry, Eastern Nazarene College, Quincy, MA 02170, USA
and
Jack W. FRAZER
Sterling Drug Inc., Malvern, PA 19355, USA

Abstract

A new method for molecular structure quantitation is described, in which both electronic and topological attributes are united. The method uses the hydrogen-suppressed skeleton to represent the structure and leads to a graph invariant index for the individual atoms and hydride groups of the molecular skeleton. An intrinsic atom value is calculated for each atom as $I = (\delta^v + 1)/\delta$, in which δ^v and δ are the counts of valence and sigma electrons of atoms in the molecular skeleton, that is, exclusive of bonds to hydrogen atoms. The electrotopological state value S_i for an atom i is defined as $S_i = I_i + \Delta I_i$, where the influence of atom j on atom i, ΔI_i, is given as $\sum(I_i - I_j)/r^2$; r is the graph separation between atoms i and j, counted as number of atoms, including i and j. The information in the electrotopological state values is revealed by examples of various types of organic structures, including chain branching and heteroatom variation. The relation of the E-state value to NMR chemical shift is demonstrated for a series of carbonyl compounds.

1. Introduction

The development of non-empirical structure descriptors has focused attention on whole molecule index generation from constituent atom or bond features. The most widely used example is the family of molecular connectivity indexes developed by Kier and Hall [1,2] from an alkane branching index by Randić [3]. These and others have been quite successful in analyzing physical and biological properties in quantitative structure–activity relationships (QSAR). The information flow in the calculation of these molecular structure indexes has been from atoms to bonds to fragments to molecules. In the calculation, however, the information at the level of atoms and bonds in molecules is subsumed into molecular indexes.

Structure specificity manifests itself in biological phenomena at the atom level in molecules. The consequence of structure change in drug design is frequently the

induction of a structural change at a position remote from the altered position. This is not commonly recognized, since attention is usually devoted to the changing substituent rather than to its effect at some distant position where the actual biological effect is influenced.

Up to now, the ability to analyze structure changes at an atom position in a molecule has largely been the realm of quantum mechanics. In refined form, these are the calculations of electron domain contours or molecular electrostatic potential maps. The problem with these analyses is the fact that they are complex contour diagrams which are not directly translatable into numerical indexes useful in equation models. The direct use of this information is excluded at this time from QSAR analyses. Some workers have used quantum chemical methods to estimate partial charges as atom indexes in QSAR. Alternative approaches not based on quantum chemical methods have been developed to estimate partial charges, including methods by Del Re [4], Gasteiger [5], and Mullay [6].

The generation of atom-level structure indexes from a chemical graph representation of a molecular structure appears to offer promise. Kier and Hall [7,8] calculated such an index from the molecular connectivity methodology to predict molecular orbital charge. By partitioning chi indexes into atom contributions, atom level indexes were calculated. This work was extended by Kier in 1987 [9] to the quantitation of the uniqueness of an atom in a molecule, and by Hall and Kier into a description of the topological equivalence of atoms [10]. Other approaches have been reported, including summed atom values to give molecular ID numbers by Randić [11], the torsion topological descriptors of Nilakantan [12], the electron-topologic approach of Bersuker [13], the topological electronic index of Kaliszan [14], a vertex topological index by Klopman [15], and development of atomic contributions for physicochemical properties by Crippen and Ghose [16].

In this paper, we reveal a new general approach to atom level indexes, an approach which incorporates both the electronic and topological character of molecular structure and is neither estimated atomic charge nor exclusively a topological descriptor.

2. Atom information fields

In this study, we recognize that every atom in a molecule is unique (except where two or more atoms may be equivalent in the sense of constitutional symmetry). This uniqueness arises from differences in the electronic and topological environment among atoms. As an example, consider the methyl carbon of propanol. It is equivalent to methyl carbons in all other propanol molecules. It is, however, different from all other atoms including other methyl carbons of other alcohols. This is true because of the differences in structure between propanol and every other molecule. Demonstrating this uniqueness requires that the particular atom in question must be compared to all other atoms in that molecule. This comparison constitutes a recognition of a field in which the atom in question resides. Each atom exists in a field composed of every

other atom in that molecule. Each field in a molecule (defined by each unique atom) is also unique. These fields, defining the uniqueness of any atom, may be considered to be information fields. It should be noted that this concept of information field is not related to information theory.

The impact of the information field upon an atom plus the intrinsic nature of that atom results in the structural attribution associated with that atom in a particular molecule. If we define this influence within a model combining electronic and topological attributes of an atom and its field (a unified attribution model), we can call the resulting index an electrotopological state of that atom.

Our objective in this study is to quantitate the information present in each information field and to assemble this into an electrotopological index associated with each atom of the molecule.

3. Quantitation of the information field

3.1. THE GENERAL MODEL

The quantitation of the effect of an information field upon any atom in a molecule may be viewed as a composite of three basic ingredients. The first is an index encoding the intrinsic topological and electronic state of any atom before account is taken of the influence of the information field upon it. The second ingredient is the quantification of the effect of the field influencing an atom. The third ingredient is the quantitation of the distance or remoteness of any part of the field from the atom under study. We can illustrate this as a field effect perturbation F, with a distance component r, operating upon an intrinsic atom value I to produce a calculated electrotopological atom state value S:

$$F(r)[I] = S.$$

3.2. THE ATOM INTRINSIC VALUE

The information to be encoded into the atom intrinsic value is both electronic and topological. The important electronic information for this model is the count of pi and lone pair electrons. Electrons occupying these orbitals are chemically most reactive and they are closely associated with the origin of the stronger, long-range intermolecular interactions. The count of these nonsigma electrons has also been shown to correlate well with the valence state electronegativity of second quantum level covalently bound atoms [1,2,9,17]. This count is equal to $\delta^v - \delta$, where δ^v is the count of valence electrons in the skeleton $Z^v - h$, and δ is the count of sigma electrons in the skeleton $\sigma - h$, where Z^v is the number of valence electrons and σ is the count of electrons in sigma orbitals.

The important topological attribute that should be encoded into the atom intrinsic value is the relative degree of mantle-atom or buried-atom status. As an example, the

methyl groups of neopentane are mantle atoms, whereas the central atom is a buried atom. To encode this information, we may use the value $1/\delta$ as a measure of the degree of mantle atom status. Terminal atoms, $\delta = 1$, tend to lie on the surface or mantle of the molecule, whereas atoms for which $\delta = 3$ or 4 are relatively buried.

It is our intention that the atom intrinsic value should be some function of $\delta^v - \delta$ and $1/\delta$. We choose the product of these terms to describe the intrinsic value I:

$$I = (\delta^v - \delta)/\delta.$$

In this equation, all of the hydrides of $C(sp^3)$ have identical values: zero. Accordingly, we scale the value of δ^v by adding one:

$$I = (\delta^v - \delta + 1)/\delta.$$

We can simplify this equation by adding one and reducing terms to obtain the equation for the atom intrinsic value:

$$I = (\delta^v + 1)/\delta. \tag{1}$$

This intuitive development leads to this definition for the atom intrinsic value, which we adopt provisionally. A list of I values for second quantum level atoms is shown in table 1. It should be emphasized that the atom intrinsic value is a nonempirical quantity.

Table 1

Intrinsic state values

Atom (group)	$[(\delta^v + 1)/\delta]$
>C<	1.25
>CH–	1.33
$-CH_2-$	1.50
>C=	1.67
$-CH_3$, =CH–, >N–	2.00
≡C–, –NH–	2.50
$=CH_2$, =N–	3.00
–O–	3.50
≡CH, $-NH_2$	4.00
=NH	5.00
≡N, –OH	6.00
=O	7.00
–F	8.00

3.3 HIGHER QUANTUM LEVEL ATOMS

The calculation of intrinsic values is based in part on the estimation of the electronegativity for 2nd quantum level atoms. This is approximated by the quantity $\delta^v - \delta$. In a series of atoms with a constant δ value such as the series F, OH, NH_2, CH_3 ($\delta = 1$), the variation in the δ^v value encodes the relative electronegativity of the group. Among the halogens, there is a constant value of δ and δ^v. To reflect adequately the differences in electronegativity among the halogens in the equation for I (eq. (1)), it is necessary to characterize in some way the relative electronegativities in the series as a function of the principal quantum number.

Accordingly, we adopt the ratio of the squares of the principal quantum number relative to the second quantum level ($N = 2$) as a modifier of the δ^v value in eq. (1). This is the ratio of surface areas relative to second quantum level atoms:

$$I = [(2/N)^2\delta^v + 1]/\delta. \tag{2}$$

The I values of several atoms and groups of higher quantum levels are shown in table 2.

Table 2

Intrinsic value I of atoms (groups) in higher quantum levels

Atom (group)	N	δ^v	δ	$[(2/N)^2\delta^v + 1]/\delta$
-Cl	3	7	1	4.111
-Br	4	7	1	2.750
-I	5	7	1	2.120
>Si<	3	4	4	0.694
>SiH-	3	4	3	0.925
>P-	3	5	3	1.074
|				
>P=	3	5	4	0.805
|				
>P<	3	5	5	0.644
-SH	3	5	1	3.222
-S-	3	6	2	1.833
=S	3	6	1	3.667

3.4. THE FIELD EFFECT ON EACH ATOM

The influence of the field upon an atom may be dissected into a summation of the interactions of all atom pairs $i \ldots j$, one of which is the atom under consideration. Each atom pair $i \ldots j$ can be viewed as defining a compartment which we can call a loge [18]. The dimension of any loge corresponds to the count of atoms in a

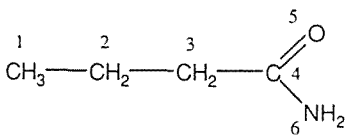
contiguous path beginning with atom i and ending with atom j. The effect of the field on an atom is thus a summation of the effect of the loges.

As an example, consider the skeleton structure of the molecule butyramide. Let us examine specifically the influence of the field upon the methyl group. The field associated with this group is composed of all atoms in the molecule. We can reckon the effect of the field by dissecting it into loges holding two atoms each. One atom in each loge is the methyl group i. The other atom is any of the remaining atoms in the molecule, atom j. The loges containing all atom $i \ldots j$ pairs are shown in table 3.

Table 3

Electrotopological state calculations for butyramide

Atom numbering:



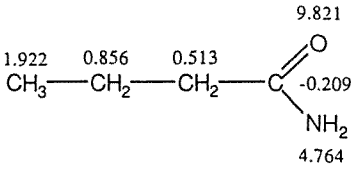
Intrinsic values:

$I(1) = 2.00 \quad I(4) = 1.67$
$I(2) = 1.50 \quad I(5) = 7.00$
$I(3) = 1.50 \quad I(6) = 4.00$

$(I_i - I_j)/r^2$ matrix

i \ j	1	2	3	4	5	6	ΔI = sum
1		0.125	0.056	0.021	− 0.200	− 0.080	− 0.078
2	− 0.125		0.000	− 0.019	− 0.344	− 0.156	− 0.644
3	− 0.056	0.000		−0.042	− 0.611	− 0.278	− 0.987
4	− 0.021	0.019	0.042		− 1.333	− 0.583	− 1.876
5	0.200	0.344	0.611	1.333		0.333	2.821
6	0.080	0.156	0.278	0.583	− 0.333		0.764
							0.000

$S_i = I_i + \Delta I_i$



To quantitate the influences of atom j on atom i within each loge, we use the intrinsic atom value, just defined. We assume that the part of the field contained within each loge has a perturbing effect on the intrinsic atom value I; this perturbation is assumed to be some function of the difference in intrinsic value I_i and I_j. Thus:

$$\Delta I = I_i - I_j. \tag{3}$$

Equation (3) is adopted as a general perturbation within each loge. The influence of atom j on atom i must decrease when the atoms are remote; thus, the size of the loge is a factor in the quantitation of the perturbation. To account for this, we modify eq. (3) with some function of r, the count of atoms in a particular loge which might also be called the graph separation.

Current work is based upon the expression $1/r^2$. If we include this term in eq. (3) and sum the influence within each loge, we arrive at an estimate of the influence of the field upon the intrinsic value of an atom:

$$\Delta I_i = \sum_{j=1}^{N} (I_i - I_j)/r^2. \tag{4}$$

The result of this influence or perturbation of I leads to an estimate of the electrotopological state S_i of any atom

$$S_i = I_i + \Delta I_i. \tag{5}$$

We refer to S_i as the electrotopological state for atom i, or simply the E-state of the atom or skeletal group.

4. Examples

4.1. SAMPLE CALCULATION

It is useful to perform a calculation on a molecule to illustrate the steps in the procedure. For this, we use the butyramide molecule. From table 1, we extract the I values for each atom. In table 3, the values are revealed in a matrix. Note that this electrotopological matrix is antisymmetric. The values of each loge contribution are obtained by adding across for each atom. This gives the field effect on each atom. The state values for each atom are then calculated and are shown at the bottom of the table.

4.2. RESULTS OF CALCULATIONS

In fig. 1 are shown the E-state values for the pentane isomers [19]. As a carbon changes from a primary through to a quaternary environment, its mantle atom status declines. The E-state values follow this structure attribute. In the same figure, a series

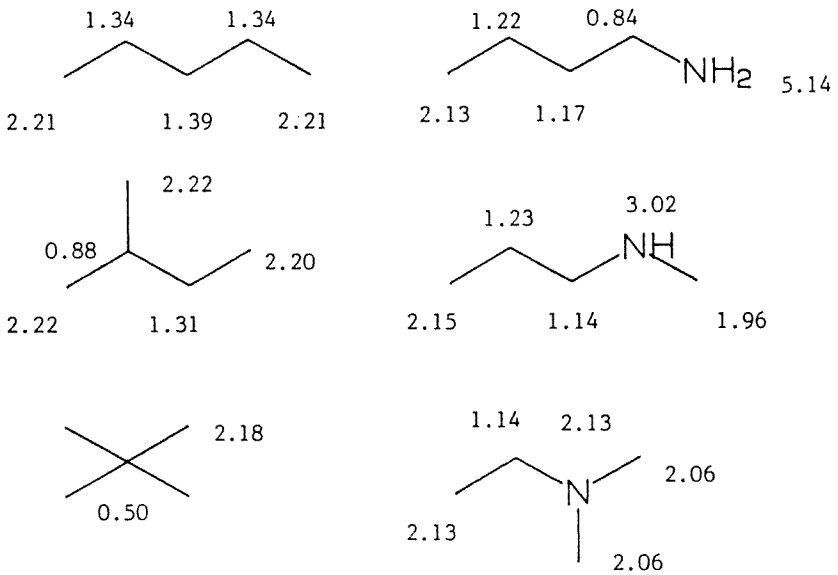


Fig. 1. Hydrogen-suppressed molecular skeletons which illustrate electrotopological state values for various topological environments.

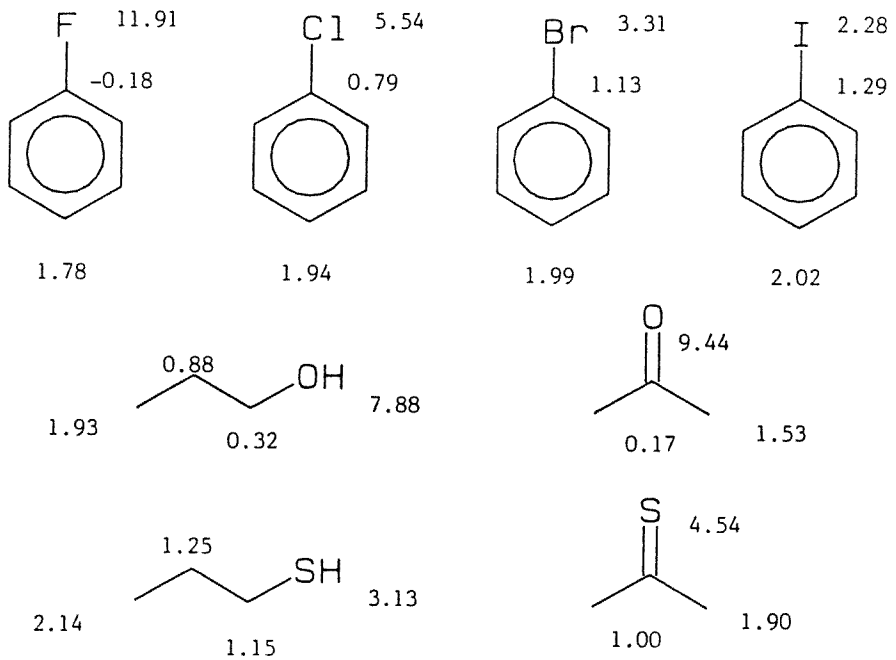


Fig. 2. Electrotopological state values for hydrogen-suppressed molecular skeletons which illustrate various heteroatomic environments.

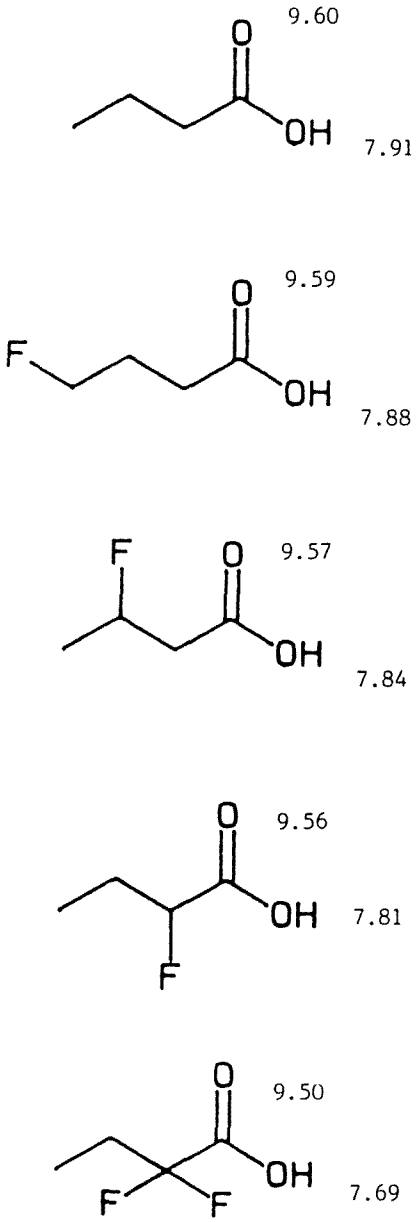


Fig. 3. Hydrogen-suppressed molecular skeletons which illustrate impact of fluorine atoms on carboxyl group atom electrotopological state values.

containing a primary, secondary and a tertiary amine is calculated. The electron-richness of the amino group leads to a higher *E*-state value relative to its corresponding methyl analog. Further, the amino groups have a greater electronegativity, leading to lower *E*-state values for contiguous atoms (groups), especially those in the alpha

and beta positions. The amine *E*-state values decrease in the direction of the tertiary derivative, as expected from its buried atom status.

In figure 2 are shown *E*-state values for the halogens and sulfur analogs of oxygen. The effect of changing the halogen on the ipso and para atoms of benzene rings is mirrored by the *E*-state values. The halogens have decreasing *E*-state values with higher quantum numbers.

Figure 3 shows *E*-state values of carbonyl groups influenced by distal and proximal electronegative atoms such as fluorine. The carbonyl atoms decline in *E*-state value under the influence of the fluorine, especially when alpha to the group or in a gem difluoro arrangement.

An examination of table 1 indicates that some skeletal groups have identical atom intrinsic values. For example, $-CH_3$, $=CH-$, $>N-$ each possess the same value $I = 2.00$. However, based on the definition of S, the *E*-state for these skeletal groups in a molecule will possess different numerical values. In the calculation of S, each of these groups enters in a different number of ways because of the differing connectivity of each in the molecular skeleton.

4.3. THE RELATION OF *E*-STATE VALUES TO NMR CHEMICAL SHIFT

The formulation of the *E*-state index for an atom is based upon the electronegativity or the pi and lone pair electron count. In addition, there is built into the index a quantitation of the topology of an atom due to its intrinsic structure and surroundings. Accordingly, we would expect that this index, the *E*-state value of an atom, would bear some relationship to a physical measurement which is dependent upon electron density and molecular topology.

Table 4

Electrotopological state values for alkyl aldehydes and ketones and ^{17}O NMR chemical shift

Obs	Compound	$S(=O)$[a]	^{17}O Delta[b]	Calc[c]	Res[d]
1	CH_3CHO	8.806	592.0	589.9	2.1
2	C_6H_5CHO	9.174	579.5	579.7	-0.2
3	i-C_3H_7CHO	9.505	574.5	570.5	4.0
4	$(CH_3)_2CO$	9.444	569.0	572.2	-3.2
5	$CH_3COC_2H_5$	9.813	557.5	562.0	-4.5
6	CH_3CO-i-C_3H_7	10.144	557.0	552.8	4.2
7	$(C_2H_5)_2CO$	10.181	547.0	551.8	4.8
8	C_2H_5CO-i-C_3H_7	10.512	543.5	542.6	0.9
9	$(i$-$C_3H_7)_2CO$	10.843	535.0	533.4	1.6

[a] Electrotopological state value for carbonyl oxygen.
[b] Measured ^{17}O NMR chemical shift [20].
[c] Chemical shift obtained from regression of experimental chemical shift on the *E*-state value for oxygen, eq. (6).
[d] res = obs – calc.

In table 4, the NMR chemical shift of ^{17}O isotopes, $\partial_{ppm}(^{17}O)$, is shown together with the E-state for the carbonyl oxygen $S(=O)$ for a series of carbonyl compounds [20]. The equation and statistics are as follows:

$$\partial_{ppm}(^{17}O) = 834.48 - 27.768\, S(O) \quad r = 0.986,\ s = 3.7,\ F = 197,\ n = 9. \qquad (6)$$

Clearly, the E-state index encodes relevant electronic and topological information of the environment of these carbonyl oxygen atoms to provide such a high quality relationship.

5. Discussion

The electrotopological state of an atom combines both an electronic and a topological description of that atom in the molecule. It is this unification of attributes which results in an index with a significant information content.

The potential for intermolecular interaction of the skeletal atom is encoded in the atom intrinsic value I through the count of pi and lone pair electrons $\delta^v - \delta$. This count is also associated with the valence state electronegativity of the skeletal atom. The important consequence of this definition arises in the formulation of the E-state value. The difference in intrinsic values ΔI contains a measure of the build-up or depletion of partial charge on the atom due to the presence in the molecule of atoms with different electronegativity. In this sense, the E-state value for an atom is directly related to the concept of atomic partial charge.

The additional feature within the E-state index is the contribution of molecular topology which enters the formalism in two ways. The atom intrinsic value I contains in the denominator the count of skeletal neighbors, a measure of the local topology of the atom. This count is the number of avenues in the skeleton over which electrons may flow from less to more polar regions. Further, in the ΔI expression, the topology of the whole molecule enters through the graph separation r. The electronegativity difference or ionicity is diminshed by the r^2 factor, where r is the count of atoms in the skeleon between pairs of atoms, counting both atoms. Hence, there is a strong proximity-related topological dependence in the E-state formulation.

Each atom in the molecule is characterized by its intrinsic state value I. The impact of every other atom, represented by eq. (4), either augments or decreases the state value, giving rise to the molecular value, the E-state index S. Even for atoms with similar or identical intrinsic values, the effects of connectedness within the molecular skeleton tend to make the final E-state values different; only topologically equivalent atoms should possess identical S values. For example, the $-CH_3$, $=CH-$, and $>N-$ groups each possess the same I value: 2.00. However, each is also characterized by a different number of neighbors in the skeleton, the δ value; 1, 2, and 3, respectively. Therefore, it is expected that these three groups will generally have different E-state values in given molecules. As shown in fig. 1, in

2-methylbutane the methyl groups have $S = 2.22$, whereas in dimethylethylamine the S value for the NH group is 2.13. The $-CH=$ group E-state is 1.88 in 1-butene and 1.64 in 1,3-butadiene.

Evidence of some of the structure information is presented in this paper. The variation of E-state values with branching agrees with usual organic intuition. Likewise, the variation of the E-state value among different heteroatoms is satisfying with respect to experience with such effects as polarity and inductive effects.

In a previous paper, the relationship between E-state index and ^{17}O NMR chemical shift was investigated for a set of alkyl ethers [21]. A high quality relation was found ($r = 0.995$, $s = 4.3$, $F = 772$, $n = 10$). In this present work, it is shown that the correlation of E-state values with ^{17}O NMR chemical shift is quite significant for a set of carbonyl compounds which includes both aldehydes and ketones. It is observed that as the E-state value increases, the chemical shift decreases. It can be observed that there is a rather general relation between the chemical shift and molecule size for this particular data set. However, compounds 2 and 4 both have three carbon atoms and six hydrogen atoms but the chemical shift differs by 10.5 ppm, whereas both these compounds are predicted well by eq. (6). Other such pairs of data points may be found in this data set. The E-state indexes have a much stronger relation to the chemical shift than does size.

It is to be emphasized that the E-state indexes are non-empirical; they are derived from counts of electrons within the hybridization model of covalent binding. They are the first atom-level indexes which combine both electronic structure and topology using the same metric, electron counts. The calculation is simple and straightforward, requiring only the element type, the hybrid state, and the connection table of atoms in a molecular skeleton [19].

The potential value of this index is apparent when the need for atom-centered, non-empirical structure descriptors is restated. There now exists the possibility to examine sub-molecular features to discover contributions toward intermolecular effects among biologically important molecules in such phenomena as protein binding and receptor interactions. Dependence upon whole-molecule indexes, valuable in their own sphere of utility, is now divested and a new paradigm is at hand to facilitate exploration of molecular mechanisms and rational design of molecules at the atomic and fragment level.

Acknowledgements

We wish to express appreciation for support of this work from Sterling Drug, Inc., Malvern, PA and Allied Signal Corporation, Engineered Materials Research, Des Plaines, IL. This work was presented at the International Chemical Congress of Pacific Basin Societies, Honolulu, December 1989.

References

[1] L.B. Kier and L.H. Hall, *Molecular Connectivity in Chemistry and Drug Research* (Academic Press, New York, 1976).
[2] L.B. Kier and L.H. Hall, *Molecular Connectivity in Structure–Activity Analysis* (Wiley, London, 1986).
[3] M. Randić, J. Amer. Chem. Soc. 97(1975)6609.
[4] G. Del Re, J. Chem. Soc. (1958)4031.
[5] J. Gasteiger and M. Marzilli, Tetrahedon 36(1980)3219.
[6] J. Mullay, J. Amer. Chem. Soc. 108(1986)1770.
[7] Ref. [1], p. 231.
[8] L.B. Kier and L.H. Hall, Tetrahedon 33(1977)1953.
[9] L.B. Kier, Quant. Struct.–Act. Relat. 65(1987)8.
[10] L.H. Hall and L.B. Kier, Quant. Struct.–Act. Relat. 9(1990)115.
[11] M. Randić, J. Chem. Inf. Comput. Sci. 24(1984)164.
[12] R. Nilekantan et al., J. Chem. Inf. Comput. Sci. 27(1987)82.
[13] I.B. Bersuker, A.S. Dimoglia and M.Yu. Gorbachov, in: *QSAR in Drug Design and Toxicology*, ed. D. Hadzi and B. Jerman-Blažić (Elsevier, Amsterdam, 1987), p. 43.
[14] R. Kaliszan, Chromatography 19 (1987).
[15] G. Klopman and C. Raychandhury, J. Comput. Chem. 9(1988)232.
[16] G. Crippen and A. Ghose, J. Comput. Chem. 7(1986)565.
[17] L.B. Kier and L.H. Hall, J. Pharm. Sci. 70(1981)583.
[18] R. Daudel, Comp. Rend. 237(1953)60.
[19] The computations on all data presented in this manuscript were performed with a new version of MOLCONN2, which will be available in the near future.
[20] C. Delseth and J.-P. Kintzinger, Helv. Chim. Acta 59(1976)466; 1411.
[21] L.B. Kier and L.H. Hall, Pharm. Res. 7(1990)801.