Pharmaceutical Research, Vol. 7, No. 8, 1990

Research Article

An Electrotopological-State Index for Atoms in Molecules

Lemont B. Kier[1,3] and Lowell H. Hall[2]

Received January 2, 1990; accepted March 5, 1990

A new method for molecular structure description is presented in which both electronic and topological characteristics are combined. The method makes use of the hydrogen-suppressed graph to represent the structure. The focus of the method is on the individual atoms and hydride groups of the molecular skeleton. An intrinsic atom value is assigned to each atom as $I = (\delta^v + 1)/\delta$, in which δ^v and δ are the counts of valence and sigma electrons of atoms associated with the molecular skeleton. The electrotopological-state value, S_i, for skeletal atom i is defined as $S_i = I_i + \Delta I_i$, for second row atoms, where the influence of atom j on atom i, ΔI_i, is given as $\Sigma(I_i - I_j)/r_{ij}^2$; r_{ij} is the graph separation between atom i and atom j, counted as the number of atoms. The characteristics of the electrotopological state values are indicated by examples of various types of organic structures, including chain lengthening, branching, heteroatoms, and unsaturation. The relation of the E-state value to NMR chemical shift is investigated for a series of alkyl ethers. The E-state oxygen value gives an excellent correlation with the ^{17}O NMR: $r = 0.993$ for 10 ethers. A biological application of the E-state values in QSAR analysis is given for the binding of barbiturates to beta-cyclodextrin.

KEY WORDS: atom index; electrotopological state; topological accessibility; quantitative structure–activity relationship (QSAR) index; intrinsic state.

INTRODUCTION

A Unified Attribution Model of Structure

There is a growing awareness that a complete definition of molecular structure is essential if we are to quantitate the influences governing measured values of physical properties or biological activities. In the past reliance on physical properties as surrogates for molecular structure quantitation was expedient since these data existed in the literature or could be measured. With the advent of all-valence-electron molecular orbital theory, tractable schemes for modeling intermolecular interactions, and contributions from a newly developing paradigm of quantitative chemical topology, it is now possible to explore molecular structure in a more definitive way.

Structural influence has been assigned to two broad categories from these studies. The first is the importance of electronic influences within molecules and between molecules. These effects have been modeled with united molecule molecular orbital calculations, interaction energy calculations, electrostatic interactions, charge-based models, and so on. These studies produce profiles of electronic influence on physical or biological effects.

The second category of structure influences is the topological attributes of molecules, called by names such as steric, spatial, bulk, volume, surface area, and so forth. These effects translate into concepts of fitting, docking, occupancy, etc., at receptors or enzyme active sites. The quantitation of this second category of structure attribute has been one of the most elusive problems to the chemist. Newly emerging contributions from topology and graph theory have been of value in this arena.

Since both these categories of structural attributions are intrinsic to a molecule, one approach seeks descriptions of structure which embrace the two in a unified model. We could think of this as a union of electronic and topological attributes of a molecule into a single quantitative description capable of modeling the measured properties or activities. We refer to this approach as a unified attribution model.

An Atom-Level Description of Structure

The development of nonempirical structure descriptors has followed the direction of whole molecule index generation from constituent atom or bond attributes. The most widely used example is the set 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 molecular properties in quantitative structure–activity relationships (QSAR). The flow of information in the calculation of these molecular structure indexes has been from atoms to bonds to fragments to molecules. The final result is a set of information-rich whole-molecule indexes, which experience has shown to be of great value in the study of physical properties and biological activities. In the general process, however, the information at the level of atoms and bonds in molecules is suppressed or subsumed into molecular indexes.

[1] Department of Medicinal Chemistry, Virginia Commonwealth University, Richmond, Virginia 23298-0581.

[2] Department of Chemistry, Eastern Nazarene College, Quincy, Massachusetts 02170.

[3] To whom correspondence should be addressed.

0724-8741/90/0800-0801$06.00/0

It is at the atom level in molecules that structure specificity manifests itself in biological phenomena. The effect of structural alteration in drug design is often a process of inducing a structural change at a position remote from the altered position. This is not commonly recognized; attention is usually devoted to the variable substituent rather than to its effect at some distant position where the actual biological effect is generated.

The ability to analyze structure changes at an atom position in a molecule has largely been the province of quantum mechanical level calculations. 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 difficult to cast into the form of QSAR equations.

The generation of atom-level structure indexes from chemical graph portrayal of molecular structure appears to offer promise. Kier and Hall (4,5) 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 (6) 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 (7). Other approaches have been reported, such as the paper by Randić on the summed atom values to give molecular ID numbers (8), the torsion topological descriptors of Nilakantan (9), the electron-topologic approach of Bersuker (10), the topological electronic index of Kaliszan (11), a vertex topological index by Klopman (12), and the development of atomic contributions for physicochemical properties by Crippen and Ghose (13).

In this paper we expand extensively upon our earlier preliminary studies and formalize a new general approach to atom-level indexes within the unified attribution model.

ATOM INFORMATION FIELDS

An initial consideration in this study is the recognition that every atom in a molecule is unique (except where two or more atoms may be mapped onto each other through a symmetry operation). This uniqueness arises from differences in the electronic and topological environment of a particular atom. As an example consider the methyl carbon of ethanol. It is equivalent to methyl carbons in all other ethanol 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 ethanol and every other molecule. This uniqueness requires that the particular atom in question must be compared to all other atoms in that molecule. This comparison uses a recognition of a field in which the atom in question resides. Each atom exists in a field of other atoms 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.

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 with a unified attribution model, we can refer to the resulting index as an electrotopological state of that atom.

Our goal is to quantitate the information resident in each information field and to incorporate this into an electrotopological index associated with each atom of the molecule.

QUANTITATION OF THE INFORMATION FIELD

The General Model

The quantitation of the effect of an information field upon any atom in a molecule requires 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 perturbing 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.

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

The Atom Intrinsic Value

The information we wish to encode 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 most chemically 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 (2,4). This count is equal to $\delta^v - \delta$, where δ^v is the count of valence electrons and δ is the count of sigma electrons in the skeleton structure.

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, while 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 tend to be less exposed to the surface.

Within the precepts of our unified attribution model, 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:

$$I \rightarrow (\delta^v - \delta)/\delta$$

If we were to use this expression, all of the hydrides of $C(sp^3)$ would have identical values, zero. Accordingly, we scale the value of δ^v by adding one.

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

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

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

A list of I values for second-quantum level atoms is shown in Table I. Treatment of higher levels is shown in Eq. 2.

Higher-Quantum Level Atoms

The calculation of I values is based in part on the estimation of the electronegativity for second-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; see Table II. In the halogens, there is a constant value of δ and δ^v ($Z^v - h$). 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 principle 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 atomic radii relative to second-quantum level atoms.

$$I = [(2/N)^2 \delta^v + 1]/\delta \quad (2)$$

The I values of several atoms and groups of higher quantum levels are shown in Table III.

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 (14). The dimension of any loge corresponds to the count of atoms in a contiguous path beginning with atom i and ending with atom j.

As an example, consider the molecule ethyl acetate. Let us examine specifically the influence of the field upon the carbonyl oxygen atom. The field associated with this atom 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 carbonyl oxygen atom, atom i. The other atom is any of the remaining atoms in the molecule, atom j. The loges containing all atom $i \ldots j$ pairs, relative to the carbonyl oxygen atom, atom 1, are shown in Table IV.

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

$$\Delta I = f(I_i - I_j) \quad (3)$$

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_{ij}, the count of atoms in a particular loge.

Current work is based upon the expression $1/r_{ij}^2$. 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.

Table I. Intrinsic-State Values

Atom (group)	δ^v	δ	$\delta^v - \delta$	I $[(\delta^v + 1)/\delta]$
>C<	4	4	0	1.25
>CH–	3	3	0	1.33
$-CH_2-$	2	2	0	1.50
$-CH_3$	1	1	0	2.00
>C=	4	3	1	1.67
=CH–	3	2	1	2.00
$=CH_2$	2	1	1	3.00
>N–	5	3	2	2.00
≡C–	4	2	2	2.50
–NH–	4	2	2	2.50
≡CH	3	1	2	4.00
$-NH_2$	3	1	2	4.00
=N–	5	2	3	3.00
=NH	4	1	3	5.00
–O–	6	2	4	3.50
≡N	5	1	4	6.00
–OH	5	1	4	6.00
=O	6	1	5	7.00
–F	7	1	6	8.00

Table II. Electronegativity and δ^v Value of Selected Isoconnective Atoms (Groups)

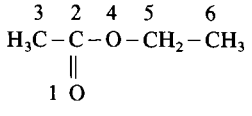
Group	δ	δ^v	$\delta^v - \delta$	X (Pauling)
–F	1	7	6	4.0
–OH	1	5	4	3.5
$-NH_2$	1	3	2	3.0
$-CH_3$	1	1	0	2.5

Table III. Intrinsic Value I of Atoms (Groups) in Higher Quantum Levels

Atom (group)	N	δ^v	δ	I, $[(2/N)^2\delta^v + 1]/\delta$
–Cl	3	7	1	4.111
–Br	4	7	1	2.750
–I	5	7	1	2.120
–SH	3	5	1	3.222
–S–	3	6	2	1.833
=S	3	6	1	3.667

Table IV. Values of $(I_i - I_j)/r_{ij}^2$ for Ethyl Acetate and Atom E-State Values

Atom numbering



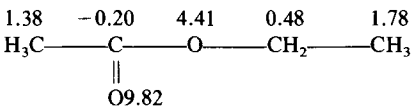
Intrinsic values

$I(1) = 7.00 \quad I(4) = 3.50$
$I(2) = 1.67 \quad I(5) = 1.50$
$I(3) = 2.00 \quad I(6) = 2.00$

$(I_i - I_j)/r_{ij}^2$ matrix

	j						
i	1	2	3	4	5	6	ΔI = sum
1		1.33	0.56	0.39	0.34	0.20	2.82
2	−1.33		−0.08	−0.46	0.02	−0.02	−1.87
3	−0.56	0.08		−0.17	0.03	0.00	−0.62
4	−0.39	0.46	0.17		0.50	0.17	0.91
5	−0.34	−0.02	−0.03	−0.50		−0.13	−1.02
6	−0.20	0.02	0.00	−0.17	0.13		−0.22
							0.00

$S_i = I_i + \Delta I_i$



$$\Delta I_i = \sum_{j=1}^{N} (I_i - I_j)/r_{ij}^2 \quad (4)$$

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

$$S = I + \Delta I \quad (5)$$

We refer to S as the electrotopological state or, simply the E-state of an atom.

EXAMPLES

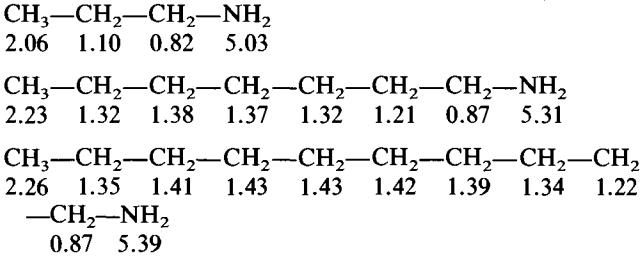
Calculation

It is useful to perform a calculation on a molecule to illustrate the steps in the procedure. For this we use the ethylacetate molecule. From Table I we extract the I values for each atom. In Table IV, the loge values are revealed in a matrix. Note that it is symmetrical across the diagonal with a sign change. The values of each ΔI are obtained by adding across for each atom. This gives the field effect on each atom. The E-state values for each atom are then calculated. These values are shown at the bottom of the table.

Results

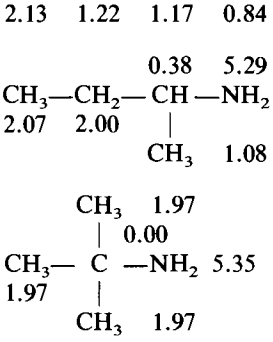
The effect of structure on state values is illustrated in several ways in the following examples.

The Effect of Chain Length



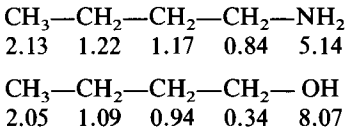
As the alkane increases by linear addition of methylene groups, the amine group gains in value. The increase approaches a constant value. The terminal methyl group exhibits the same behavior. This calculated effect mirrors the inductive effect and also the increase in topological freedom or diminution in steric crowding.

Effect Due to Branching



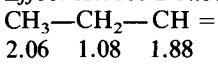
Increased branching adjacent to the amine group results in an increased value at the amine group. This mirrors the inductive effect ranking of tertiary > secondary > primary alkyl groups. The small positive value of the alpha carbon models both the electron loss and the topological inaccessibility or buried status due to steric crowding.

Effect Due to Heteroatom Change



The gain in state value at the hydroxyl group is greater than that experienced at the amine group reflecting the influence of the relative electronegativities. The trend in the E-state values from methyl to amino to hydroxyl is the same as the trend in the valence-state electronegativities of the carbon, nitrogen, and oxygen atoms in the valence states in these groups.

Effect Across Double Bonds



The sp^2 carbon atom across the double bond from the alkyl group gains in E-state value relative to the other sp^2 carbon

atom. It is a phenomenon reflecting the mantle atom status of one carbon sp^2 atom and the relatively buried status of the other member of the double bond.

THE MEANING OF THE E-STATE INDEX

The E-state value is derived from the value of the electronegativity, distributed over an atom according to its bonding degree to nonhydrogen atoms. This is encoded into the expression $(\delta^v - \delta)/\delta$ used in the derivation of the intrinsic state value of an atom. Further refinements of the equation account for differences among hydrides of carbon and the principal quantum level of the atom.

The intrinsic electrotopological state, I, of the atom is perturbed by the presence of each of the other atoms in its field or in the molecule. This perturbation is reckoned in part from the electronegativity differences.

The resulting electrotopological state (E-state) of any atom is a numerical value depicting the accessibility of that atom to interaction across space with some reference atom or group. This calculated accessibility includes both electronegativity and the topological environment, that is, being a buried atom or being an atom on the periphery of the molecule, referred to as a mantle atom.

This accessibility may be thought of as being a susceptibility or a probability of interaction of an atom under study with atoms in some other molecule. In the case of the atoms with large E-state values, the hypothetical interaction prediction model could involve a hydrogen bond. Thus the strength of the hydrogen bond (as an acceptor) parallels the series $-F > =O > -OH > -N< > -NH_2$.

The E-state values also parallel the extent to which an atom is on the exposed surface of the molecule and is buried within the skeleton of the molecule. For the hydrides of nitrogen, this trend accounts for lowered mantle atom status on going from primary to secondary to tertiary, thus the ranking $-NH_2 > -NH- > -N<$.

At the lower end of the E-state scale of atoms the hydrocarbon groups appear. Hydrogen bonds or dipolar forces are not prominent here. We may invoke a model of dispersion bonding as a possible interaction force corresponding to the low-order E-state values computed for these atoms. This is consistent with the presence of one- or two-pi electrons and/or the relative mantle atom status of the molecule.

With this proposed explanation of the possible significance of the E-state values at hand, it should be possible to use these atom descriptors in the QSAR analyses of many biological phenomena.

THE RELATION OF E-STATE VALUES TO NMR CHEMICAL SHIFT

The formulation of the intrinsic value I contains the electronegativity of the atom in terms of the Kier–Hall valence-state electronegativity, $\delta^v - \delta$ (2,4). The formalism for obtaining the E-state value, S_i, for atom i contains the electronegativity difference between atom i and the other atoms in the molecule; this difference relates to the ionicity. Further, as indicated above, the E-state value also directly contains the effect of molecular topology through the use of the count of skeletal neighbors in the denominator of Eq. 2. Thus, it is expected that E-state values may have a relation to effects which are dependent on electron density as influenced by molecular topology.

We have examined a set of 10 alkyl ethers and their measured ^{17}O NMR chemical shifts (15). The E-state value for the ether oxygen, $S(-O-)$, gives a very high correlation with the ^{17}O chemical shift, ∂, as follows:

$$\partial = -441.65 + 92.564*S(-O-)$$
$$r = 0.995, \quad s = 4.3, \quad F = 772, \quad n = 10 \quad (6)$$

The standard deviation allows for the chemical shifts to be well predicted as well as to be correctly ranked by the E-state value $S(-O-)$ as shown in Table V.

THE BINDING OF BARBITURATES TO CYCLODEXTRIN

In an effort to explore the significance of the E-state values in biological phenomena, we have undertaken studies in which there is an interaction between members of a series of compounds and a common receptive molecule. Such is the case in the analysis of the binding of barbiturates to cyclodextrin, reported by Uekama *et al.* (17). Lopata *et al.* (18) have analyzed Uekama's data with indicator variables and molar refraction. Kier has analyzed the same data using the kappa shape indexes (19).

Both of these studies are built around whole-molecule descriptors, leading to general relationships as a function of the varying side chains on the barbiturates. There is no information generated about the specific atoms in the barbiturates which may be involved in the cyclodextrin binding. Using the E-state values of the ring atoms of the barbiturates, we can explore the possibility of a specific atom involvement and we can create a QSAR model to describe this involvement.

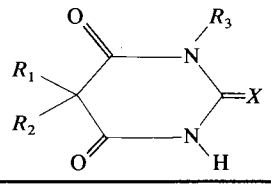
In this analysis, we studied both alpha-cyclodextrin and beta-cyclodextrin stability constants, log K_a and log K_b, with a series of barbiturates shown in Table VI. The E-state values for the ring atoms were calculated as described above. Correlations were sought between the log K of each form of

Table V. Electrotopological-State Values for Alkyl Ethers and ^{17}O NMR Chemical Shift

Obs	Compound	$S(-O-)^a$	^{17}O delta^b	Calc^c	Res^d
1	Dimethylether	4.20	−52.2	−52.9	0.7
2	Ethylmethyl−	4.54	−22.5	−21.4	−1.1
3	i-Propylmethyl−	4.75	−2.0	−2.0	−0.0
4	t-Butylmethyl−	4.94	8.5	15.6	−7.1
5	Diethyl−	4.83	6.5	5.4	1.1
6	i-Propylethyl−	5.04	28.0	24.9	3.1
7	t-Butylethyl−	5.23	40.5	42.4	−1.9
8	Di-i-propyl−	5.25	52.5	44.3	8.2
9	t-Butyl-i-propyl−	5.44	62.5	61.9	0.6
10	Di-t-butyl−	5.63	76.0	79.5	−3.5

[a] Electrotopological state value for ether oxygen.
[b] Measured ^{17}O NMR chemical shift (15).
[c] Chemical shift obtained from regression of experiment chemical shift on the E-state value for oxygen, Eq. (6).
[d] Res = obs − calc.

Table VI. Binding Constants K_b for Barbiturate Binding to Beta-Cyclodextrin and Electrotopological-State Value for Carbonyl Oxygen



No.	R_1	R_2	R_3	X	E-state value $S(=O)$	Log K_b Obs[a]	Log K_b Calc[b]
1	$-CH_2CH_2CH_3$	$-CH_2CH_3$	H	O	11.565	2.114	2.282
2	$-CH_2CH_2CH_2CH_3$	$-CH_2CH_3$	H	O	11.681	2.681	2.737
3	$-CH_2CH_2CH_2CH_2CH_3$	$-CH_2CH_3$	H	O	11.770	3.114	3.010
4	$-CH_2CH_2CH_2CH_2CH_2CH_3$	$-CH_2CH_3$	H	O	11.839	3.456	3.175
5	$-CH(CH_3)-CH_2CH_2CH_3$	$-CH_2CH_3$	H	O	11.915	3.196	3.311
6	$-CH_2CH_2CH(CH_3)_2$	$-CH_2CH_3$	H	O	11.788	3.243	3.057
7	-phenyl	$-CH_2CH_3$	H	O	11.976	3.270	3.385
8	-phenyl	$-CH_2CH_3$	CH_3	O	12.344	3.220	3.165
9	-cyclohex-3-enyl	$-CH_3$	CH_3	O	12.187	3.185	3.398
10	$-CH_2CH_3$	$-CH_2CH_3$	H	S	11.566	2.477	2.286
11	$-CH_2CH_2CH_3$	$-CH_2CH_3$	H	S	11.725	2.732	2.880
12	$-CH_2CH_2CH_2CH_3$	$-CH_2CH_3$	H	S	11.841	2.839	3.179
13	$-CH_2CH_2CH_2CH_2CH_3$	$-CH_2CH_3$	H	S	11.930	3.324	3.332
14	$-CH_2CH_2CH_2CH_2CH_2CH_3$	$-CH_2CH_3$	H	S	11.999	3.684	3.404
15	$-CH(CH_3)-CH_2CH_2CH_3$	$-CH_2CH_3$	H	S	12.052	3.380	3.433
16	-phenyl	$-CH_2CH_3$	H	S	12.136	3.549	3.429

[a] From Ref. 17.
[b] From Eq. (7).

cyclodextrin and the E-state values of atoms associated with the barbiturate ring.

The study on the alpha-cyclodextrin produced no significant correlation with any barbiturate E-state value or pair of values. In contrast, a significant correlation was found in the case of the beta-cyclodextrin. (The heptyl derivative compound was a significant outlier and was dropped from further regression analysis; the reasoning for deletion is that its length may preclude its satisfactory fit in the cyclodextrin cavity, which is about 7.8 Å deep.)

The equation model is a quadratic expression in the E-state value for one carbonyl oxygen atom, symbolized as S(=O), shown in Table VI. This oxygen atom is next to the R_3N.

$$\log K_b = 101.84 * S(=O) - 4.21 * S(=O)^2 - 612.16$$
$$r = 0.90, \quad s = 0.20, \quad F = 28, \quad n = 16 \qquad (7)$$

The results indicate that the electrotopological state of the carbonyl oxygen atom is defining certain attributes of this atom which are highly influential in the binding of these molecules to beta-cyclodextrin. We can interpret this result as an indication that the carbonyl oxygen atom is participating in the binding to beta-cyclodextrin along with the barbiturate side chain. The measured difference in binding is apparently a strong function of the electrotopological state of this atom.

In contrast, the alpha-cyclodextrin is indifferent to the E-state of the ring atoms and is probably binding via the side chain. Thus we see no effect due to the barbiturate ring atoms.

These results coincide exactly with the results reported by Lopata *et al.* (18). They concluded that beta-cyclodextrin is binding only the side chain, while beta-cyclodextrin is binding the side chain plus a part of the ring. Using our E-state calculations, we can focus attention on a particular atom of the barbiturates and show its structural correlation with the measured property.

DISCUSSION AND CONCLUSION

The electrotopological state of an atom combines both an electronic and a topological description of that atom in the molecule. It is this combination of attributes which produces the fruitful combination which is the E-state index.

The valence state 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 closely 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 of intrinsic values, ΔI, encodes both structural and topological attributes which arise from electronegativity differences and topological connectivity. In this sense the E-state value for an atom is in part related to the concept of atomic partial charge.

The additional feature of the E-state 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 the local topology of the atom. This count is the number of avenues in the skeleton over which electrons may have influence from less

to more polar regions. Further, in the ΔI relation the topology of the whole molecule enters through the graph distance, r. The electronegativity difference or ionicity is diminished by the r^2, where r is the count of atoms in the skeleton between and including pairs of atoms. Hence there is a strong topological dependence in the E-state formulation.

Evidence of significant structure information is presented in this paper in several ways. The variation of E-state values with alkyl-chain lengthening and branching agrees with usual organic intuition. Likewise, the variation of the E-state value as different heteroatoms are considered is satisfying with respect to experience with related compounds in such effects as polarity and inductive effects.

It is shown that the correlation of E-state values with ^{17}O NMR chemical shift is quite significant. It is observed that as the E-state value increases, so also does the chemical shift. Further, as the immediate environment of the ether oxygen becomes more crowded, as in the t-butyl substituents, both the chemical shift and the E-state value continue to covary significantly.

As an example of biological QSAR, the binding of barbiturates to beta-cyclodextrin was investigated. It is found the the E-state value for the carbonyl oxygen is closely related to the binding constant. The relation is quadratic, indicating that there is an optimum molecular geometry for binding. The optimum E-state value is $S(=O)_{opt} = 12.095$ and the optimum binding constant, according to this relation for these compounds in Eq. (7), is $\log K_{b,opt} = 3.718$.

This work presents the development of an index which combines both the electronic and the topological attributes of skeletal atoms. This is the first atomic-level topological index which combines these two important aspects of molecular structure. The index is computed in a straightforward manner.[4] The E-state index is easy to follow in derivation, significant in information scope, and potentially of great value in structure–activity analysis.

ACKNOWLEDGMENTS

We wish to express appreciation for valuable discussions with Jack W. Frazer and for support of this work from Sterling Drug, Inc., Malvern, PA, and Allied Signal Corporation, Engineered Materials Research, Desplaines, IL.

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,* John Wiley, London (1986).
3. M. Randić. *J. Am. Chem. Soc.* 97:6609–6615 (1975).
4. L. B. Kier and L. H. Hall. *J. Pharm. Sci.* 70:583–589 (1981).
5. L. B. Kier and L. H. Hall. *Tetrahedron* 33:1953–1957 (1977).
6. L. B. Kier. *Quant. Struct-Act. Relat.* 6:8–12 (1987).
7. L. H. Hall and L. B. Kier. *Quant. Struct.-Act. Relat.* (in press).
8. M. Randić. *J. Chem. Inf. Comput. Sci.* 24:164–168 (1984).
9. R. Nilekantan *et al. J. Chem. Inf. Comput. Sci.* 27:82–85 (1987).
10. I. B. Bersuker, A. S. Dimoglia, and M. Yu. Gorbachov. In D. Hadzi and B. Jerman-Blazic (eds.), *QSAR in Drug Design and Toxicology,* Elsevier, Amsterdam, 1987, pp. 43–48.
11. R. Kaliszan. *Chromatography* 29:19–29 (1987).
12. G. Klopman and C. Raychandhury. *J. Comp. Chem.* 9:232–243 (1988).
13. A. K. Ghose and G. M. Crippen. *J. Chem. Inf. Comput. Sci.* 27:21–35 (1987).
14. R. Daudel. *Compt. Rendu.* 237:60–65 (1953).
15. C. Delseth and J.-P. Kintzinger. *Helv. Chim. Acta.* 59:466–472 (1976); 59:1411–1419 (1976).
16. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. P. Stewart. *J. Am. Chem. Soc.* 107:3902–3908 (1985).
17. K. Uekama, F. Hirayama, S. Nasu, N. Matsuo, and T. Irie. *Chem. Pharm. Bull.* 26:3477–3482 (1978).
18. A. Lopata, F. Darvas, A. Stadler-Szoke, and J. Szejtli. *J. Pharm. Sci.* 74:211–213 (1985).
19. L. B. Kier. *Quant. Struct.-Act. Relat.* 5:1–7 (1986).

[4] The computations on all data presented in this paper were performed with a new version of MOLCONN2 which will be available in the near-future.