Tetrahedron Vol. 36, pp. 759 to 768

0040-4020/80/0315-0759/$02.00/0

STERIC EFFECTS—II

RELATIONSHIP BETWEEN TOPOLOGY AND THE STERIC PARAMETER. E'_S—TOPOLOGY AS A TOOL FOR THE CORRELATION AND PREDICTION OF STERIC EFFECTS

ANNICK PANAYE,[1a] JOHN ANTHONY MACPHEE[1b] and JACQUES-EMILE DUBOIS*

Institut de Topologie et de Dynamique des Systèmes de l'Université Paris VII, associé au C.N.R.S., 1, rue Guy de la Brosse, 75005 Paris, France

(*Received in UK 2 April 1979*)

Abstract—The steric effect of alkyl groups as characterized by the revised Taft E'_S parameter is analysed using an approach based on the DARC topological system and its PELCO correlation method. This approach involves an analysis of the systemativ variation of E'_S in a topological dequencing of alkyl groups and shows the existence of three regions of distinct behaviour: R I, a "normal" behaviour region (*ca* 6 E'_S units) in which the contribution of the introduction of successive Me groups to the overall steric effect *increases monotonically* (groups with 1 to 7 carbons); R II, a region in which a "levelling" effect is observed, i.e. the contribution diminishes and becomes nil (groups with 8 and 9 carbons); and R III, where this contribution changes sign, "inversion" effect (groups with 10 carbons).

Using a series of successive approximations, topological models are developed and tested. The conditions under which the topology may be used to represent the topography (i.e. the real 3-dimensional structure) are considered. The correlation of existing E'_S values and the reliable prediction of experimentally unavailable steric effects are direct consequences of this treatment.

In the preceding article[2] of this series we have developed an homogeneous scale of steric effects based on the Taft–Ingold hypothesis and an exact specification of the defining reaction (acid-catalysed esterification of carboxylic acids in MeOH, at 40°). This revised Taft scale, E'_S, permits, for the first time, a sound analysis of the behaviour of the steric effect with increasing substitution. Insofar as this behaviour is not trivial its understanding presents an interesting and, we feel, fruitful problem in chemistry.

The establishment of relationships or, in a more strict sense, correlations, as well as the possibility of prediction often constitutes the first steps in the comprehension of a phenomenon. In order to investigate the relationship between structure and the steric parameter E'_S, we have used a topological approach based on the DARC system[3] in which the evolution of a property within a population of ordered compounds is examined to bring out the diverse tendencies of the data and the possible existence of anomalous behaviour. It yields excellent structural correlations of predictive value and requires no *a priori* hypotheses concerning the nature of the phenomenon under consideration. In this approach groups are viewed as a set of individual sites which may easily be treated within a graph theory framework.

Formal sequencing of alkyl groups—Augmentation, levelling and inversion of their E'_S values

Previous work by Dubois *et al.* has shown how it is possible to establish purely formal relationships between the structures of alkyl groups and properties.[4] The nature and utility of these relationships depends on the ordering which is imposed on these structures. The classical notion of homology is logically contained in this study of formal populations. Furthermore, although usually associated with linear structures, the term may be generalized to include branching.

Considering that alkyl groups may be derived from one another by the formal replacement of an H by a Me group, it is possible to interrelate them by a formal link symbolising the replacement. The structures may then be included in a "hyperstructure"[5,6] which brings out the formal relationships which exist among them (Fig. 1). This representation of a population in graphical form, where the nodes correspond to compounds and the links to relationships between the compounds, has previously been used to describe relationships of various kinds (reactions,[7] isomerism,[8] configuration,[9] conformation[10]) including purely formal ones.[6,11]

The topological description of an alkyl group, on which the hyperstructure is based, represents a loss of information with respect to the real three dimensional structure. For the simpler groups the existence of standard geometry and free rotations about bonds allows the topological description to represent the real molecule faithfully. On the other hand, when there are interactions which give rise to privileged conformations and/or deformations, additional information (experimental or calculated) is required to interrelate the topology and the three dimensional structure. For this reason, the hyperstructure constitutes only a first order analysis which is subject to refinement. It permits us to determine general tendencies and to test the reliability of experimental data.

We shall use this formal sequencing of alkyl groups to analyse the influence of introducing a Me into an alkyl group on its steric parameter E'_S. The hyperstructure, shown in Fig. 1, is equivalent to a labelled and ordered conversion graph of the known alkyl groups, homomorphous to the group $t\text{-}Bu_3C\text{–}$. From it, it is evident that groups with the same number of C atoms have steric effects of the same order of magnitude. For families of isomers the order of increasing steric hindrance is: primary < secondary < tertiary. This is true for all isomer

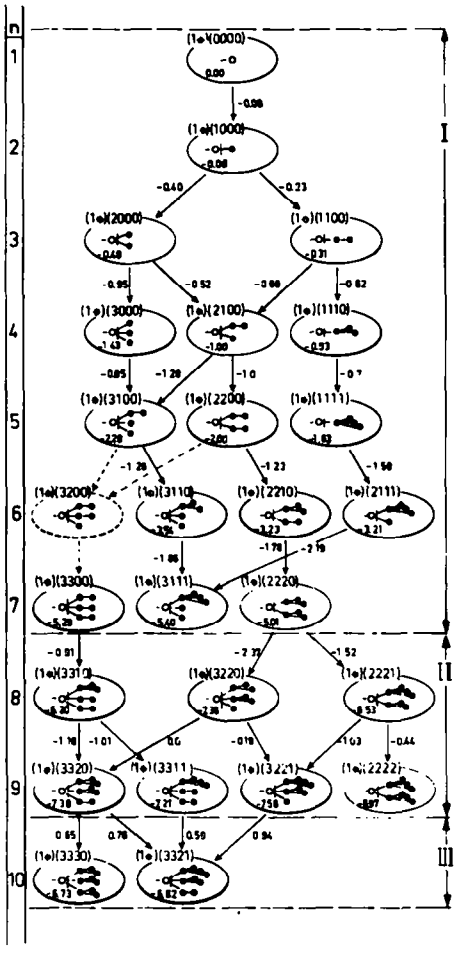


Fig. 1. Hyperstructure (or formal evolution graph) representing the evolution of the steric parameter E'_S of alkyl groups homomorphous to the group t-Bu_3C– as a function of substitution. Contained in each oval (i.e. node of the graph) are, from top to bottom, the numerical descriptor (DEL) of the alkyl group (see Fig. 3), its relevant structure and E'_S value. Over each arrow the E'_S difference between successive groups is given. This hyperstructure brings out the existence of three distinct regions: R I—successive methyl groups produce monotonically increasing contributions to the overall E'_S value, R II—these contributions become smaller and even nil (*levelling* effect), R III—a definite *inversion* effect, i.e. a decrease in the absolute value of the overall steric effect.

families from n-Pr (3 carbons) to t-$BuMe_2C$– (7 carbons). By indicating on each link of the hyperstructure the difference between the steric parameters of the groups which are derived from one another by monosubstitution one may visualize the variation of E'_S due to *the successive introduction of* Me *groups*. A simple examination of Fig. 1 shows three regions which are, in fact, distinguished by the effect of the successive introduction of Me groups: R I—up to and including groups with 7 C atoms in which the contribution of successive Me groups increases monotonically; R II—an intermediate region (groups with 8 and 9 C atoms) in which this contribution diminishes and becomes nil (except for the group i-Pr_2MeC–); and R III—an extreme region where this contribution changes sign (groups with 10 C atoms).

This behaviour is more strikingly portrayed in Fig. 2 in which the ordinate $\delta E'_S$ $(n \rightarrow n+1)$ represents the

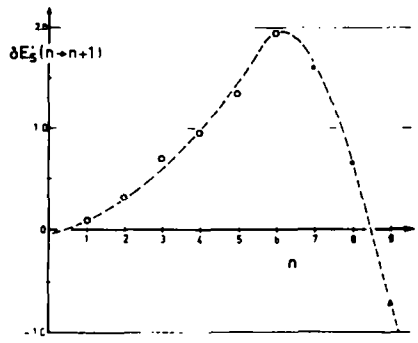


Fig. 2. Variation of E'_S of alkyl groups with changing level of substitution, $\delta E'_S$ $(n \rightarrow n+1)$ versus the number of carbon atoms—n. ○, R I; ●, R II; ▲, R III.

average variation of steric effect in alkyl groups and, the abscissa the number of C atoms in the corresponding groups. The monotonic (and relatively linear) contribution of successive C atoms in Region I is brought out as well as the levelling effect of Region II and the inversion of sign of Region III. It is to be noted that Δ represents, for a given level of substitution (n), *the average contributions to the overall steric effect of atoms possessing different environments* and as such constitutes a semi-quantitative point of view. In the following article we will come back to this problem in mechanistic terms where we will see the importance of conformational preference.

It would be futile to seek a single explanation for the augmentation (R I), levelling (R II), and inversion (R III) of the steric parameter. These domains should be considered separately since the effects which are operating are surely distinct. However, a complete data analysis does constitute a logical starting point for our topological treatment.

Darc topological analysis of the E'_S *parameter*

(a) *Correlation in terms of groups and sites.* In order to relate compound structure to a particular phenomenon (reaction rates, spectroscopic data, etc.) it is conventional to make use of correlations in terms of group properties (LFER);[12] correlations in terms of sites,[4] less generally used, are also possible. In the former case one may use, for example, steric and/or electronic parameters to reflect substituent character. The applicability of substituent parameters depends on the existence of similar interaction mechanisms in different reaction systems. Their use is therefore limited to the correlation of properties of closely related systems.

For correlations in terms of sites, however, the importance of each site (atom or bond) is evaluated for each phenomenon studied using only the data corresponding to the system under consideration and the topology. Our aim has been to use a topological correlation technique (in terms of sites) for the E'_S steric parameters of alkyl groups. We have shown, to date, how a rather large body of experimental data (reactivity,[13] spectroscopic,[14] pharmacodynamic,[15] etc) can be correlated in terms of topology alone. In the absence of important molecular distortion and/or conformational change the topography may be adequately represented by the topology.

To take into account the influence of alkyl groups on molecular properties, a simplified *non-bonded interaction*

model is proposed in which each site is associated with an intrinsic contribution together with interactions between non-bonded atoms. In this approach we have sought not to minimize the total number of interactions but rather to minimize the number of numerically different interactions, by taking advantage of symmetry. To carry out a treatment in terms of sites requires, *a priori*, that the sites be ordered. Before considering the correlation results we will briefly discuss the DARC/PELCO correlation[4] technique and the DARC order which are to be used here.

(b) *Concept of an environment which is limited, concentric and ordered* (ELCO). The DARC topological system is based on the concept of an Environment $\mathscr{E}$ conceived in terms of concentric modules, in which atoms and bonds are unambiguously specified. This environmental approach makes use of both graph theory[16] and the notion of order.

In the DARC approach any chemical structure may be represented by a chromatic[a] graph[3] consisting of a set of nodes (any atom other than hydrogen) and links (bonds) which join these nodes. In this graph we choose an origin called the focus FO whose environment $\mathscr{E}$ is organized in concentric layers of atoms, termed alternatively A and B. Two successive layers A and B constitute an environment limited in B (E_B). The environments E_B are numbered according to their proximity to the focus: E_B^1, $E_B^2, \ldots, E_B^n$ (Fig. 3).

The site labels A_i and B_{ij} are obtained by means of topological ordering criteria applied to an environment which is limited and concentric. The propagation[6] of this Environment, which is Limited, Concentric and Ordered (ELCO),[17] over the total environment $\mathscr{E}$ permits the indexing of each site of the environment unambiguously. *The molecule is then represented by a totally ordered graph.* A "connexe" matrix[3] corresponding to this graph can be used for purpose of description. This matrix is broken up into smaller units which contain, separately, the topological and chromatic information on the nature of the atoms and bonds. The numerical Descriptor of a compound is then obtained by concatenation of all the partial Descriptors of the Environment which is Limited, concentric and ordered (DEL). [b]This DEL consists of the Descriptor of Existence (DEX) which indicates the topology and the chromatic descriptors (DLI: Descriptor of LInks and DNA: Descriptor of the Nature of Atoms) to which it is possible to annex, if required, descriptors specifying stereochemistry, chirality,[19] and conformation.[20] Alkyl groups, with which we are mainly concerned, are sufficiently described by the DEX, which is expressed as $\Sigma A_i\ \Sigma B_{i1}\ \Sigma B_{i2}\ \Sigma B_{i3}$ (Fig. 4).

Topology/E_S correlation for alkyl groups-data treatment with independent site perturbations and interaction terms

[a]The chromatism includes all the information that must be added to the existence graph in order to describe the planar structural formula of a compound. The existence graph is analogous to the saturated aliphatic hydrocarbon with the same number of heavy atoms, carbons and heteroatoms as the compound under consideration.

[b]The DEL is an alphanumeric descriptor containing, in sequence, information on the planar structural formula, the stereochemistry, and the geometry within the framework of the general DARC language. It is as precise as the problem at hand requires. It is worth noting that, in the absence of such a general language, the need for treating topological and geometrical data has frequently led to partial solutions to the problem.[18]

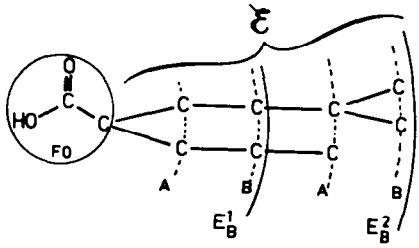
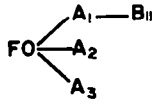


Fig. 3. Concentric organization of the environment $\mathscr{E}$ in the DARC system. For acids the focus (FO) chosen is HOCOC ═. The environment of this FO is organized in concentric layers. Two successive layers constitute an Environment E_B limited in B.

In the general DARC/PELCO[21] method the "perturbation" of a property resulting from the formal substitution of a compound in a series is estimated. This perturbation is associated with the site introduced even though, in reality, it is transmitted and redistributed over the entire molecule. The set of perturbation terms associated with the different sites of the environment is calculated by recursion. In simple cases the perturbations are additive. The most general equation, however, contains interaction terms which are to be expected in the present case since the steric parameter E_S' is not additive in nature.[22,23]

In effect, without taking into consideration all interactions between non-bonded atoms, the simultaneous presence of two sites which are independent according to the DARC generation law sometimes requires the consideration of interaction terms. For example, the exis-



tence of the site B_{11} implies the existence of the site A_1, while the sites A_2 and A_3 can exist independently of B_{11}, with non-negligible $A_2 - B_{11}$ and $A_3 - B_{11}$ interactions.

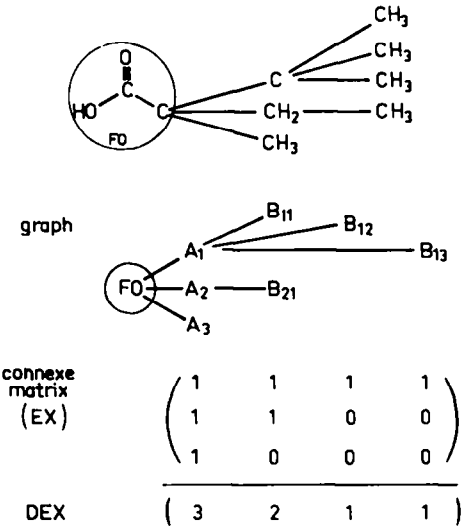


Fig. 4. Definition of the DEX. The Descriptor (DEX) is obtained from the sum of the elements of the connexe matrix of the graph representing the group. $DEX = \Sigma A_i\ \Sigma B_{i1}\ \Sigma B_{i2}\ \Sigma B_{i3}$. For alkyl groups where links and nodes are identical, DEX ═ DEL, i.e. general alphanumeric descriptor of structures.

A preliminary correlation on the overall data using both positive and negative perturbation and interaction terms in order to account for the augmentation, levelling and inversion effects leads to an acceptable result ($r = 0.9937$, $\psi = 0.17$, $s = 0.22$).[24] This brings out the significant departure from additivity of the site perturbations (Fig. 5a). However, a number of rather large discrepancies between experimental and calculated values prevents its use in the reliable prediction of new steric constants (Table 1).

In order to have an approach with greater prediction capacity we have restricted the population to the groups homomorphous to the group t-$BuMe_2C$-. In this regard we have tried to take into consideration all non-negligible interactions. It appeared at the outset that all associations which are independent according to the DARC generation laws require interactions. The correlation thus obtained (Corr. 2) is excellent ($r = 0.999994$, $\psi = 0.0068$, $s = 0.00034$) but it does contain 8 parameters for 13 data points (Fig. 5b, Table 1).

It is, of course, trivial to consider the creation of new interactions to treat each new compound in a correlation; we have, therefore, looked for a simpler model in which the interactions follow some general law. This approach appears reasonable when one considers the equivalence of certain interactions. We have chosen a simple, yet satisfying, model rather than the basic PELCO treatment in view of the restricted nature of the population to be considered. The interactions between non-bonded atoms, considered here, lead to a more topographical model; at this point, however, their justification in physical terms is not sought after.

Topology E_S' correlations—site contribution model with a priori *interactions*

(a) *$\mathscr{C}/E_S'$ Correlation for alkyl groups in the first environment* (E_B'). In the simple model considered here we have divided the perturbation due to the introduction of a site into an effective contribution and implicit inter-

Table 1. Topological correlation $\mathscr{C}/E_S'$ values of alkyl groups restricted to an environment with two concentric layers, A and B—data treatment with independent site perturbations and interaction terms

GROUP	DEL	N° of C	E_S' exp	CORRELATION 1		CORRELATION 2	
				E_S' cal	Δ	E_S' cal	Δ
CH_3-	(1*)(0000)	1	0.00	0.00	0.00	0.00	0.00
$MeCH_2$-	(1*)(1000)	2	-0.08	+0.35	0.43	-0.09	0.01
$EtCH_2$-	(1*)(1100)	3	-0.31	-0.20	0.11	-0.31	0.00
Me_2CH-	(1*)(2000)		-0.48	-0.42	0.06	-0.47	0.01
i-$PrCH_2$-	(1*)(1110)	4	-0.93	-1.19	0.26	-0.93	0.00
EtMeCH-	(1*)(2100)		-1.00	-0.97	0.03	-1.00	0.00
Me_3C-	(1*)(3000)		-1.43	-1.92	0.49	-1.43	0.00
t-$BuCH_2$-	(1*)(1111)	5	-1.63	-1.91	0.28	-1.63	0.00
i-PrMeCH-	(1*)(2110)					-1.94	-
Et_2CH-	(1*)(2200)		-2.00	-2.33	0.33		
$EtMe_2C$-	(1*)(3100)		-2.28	-2.47	0.19	-2.28	0.00
t-BuMeCH-	(1*)(2111)	6	-3.21	-3.09	0.12	-3.22	0.01
i-PrEtCH-	(1*)(2210)		-3.23	-3.33	0.10		
i-$PrMe_2C$-	(1*)(3110)		-3.54	-3.46	0.08	-3.54	0.00
Et_2MeC-	(1*)(3200)						
t-BuEtCH-	(1*)(2211)	7					
i-Pr_2CH-	(1*)(2220)		-5.01	-5.23	0.22		
Et_3C-	(1*)(3300)		-5.29	-4.91	0.38		
i-PrEtMeC-	(1*)(3210)						
t-$BuMe_2C$-	(1*)(3111)		-5.40	-5.00	0.40	-5.40	0.00‡
i-$PrEt_2C$-	(1*)(3310)	8	-6.20	-5.91	0.29		
t-BuEtMeC-	(1*)(3211)						
t-BuPr^{i}CH-	(1*)(2221)		-6.53	-6.09	0.44		
i-Pr_2MeC-	(1*)(3220)		-7.38	-6.73	0.65		
t-Bu_2CH-	(1*)(2222)	9	-6.97	-6.97	0.00‡		
t-$BuEt_2C$-	(1*)(3311)		-7.21	-7.44	0.23		
i-Pr_2EtC-	(1*)(3320)		-7.38	-7.81	0.43		
t-BuPr^{i}MeC-	(1*)(3221)		-7.56	-8.00	0.44		
t-BuPr^{i}EtC-	(1*)(3321)	10	-6.62	-6.62	0.00‡		
i-Pr_3C-	(1*)(3330)		-6.73	-6.73	0.00‡		
t-Bu_2MeC-	(1*)(3222)						

‡ For these cases Δ = 0 is trivial since last site perturbation is valuated from single datum.

Correlation 1 covers complete experimental population, but has slight predictive capacity.

Correlation 2 covers population restricted to compounds in the trace of t-$BuMe_2C$; it is precise but requires 8 parameters for only 13 data points.

Correlation 1 $r = 0.9937$, $\psi = 0.17$, $s = 0.22$, $n = 25$

Correlation 2 $r = 0.999994$, $\psi = 0.0068$, $s = 0.00034$, $n = 13$

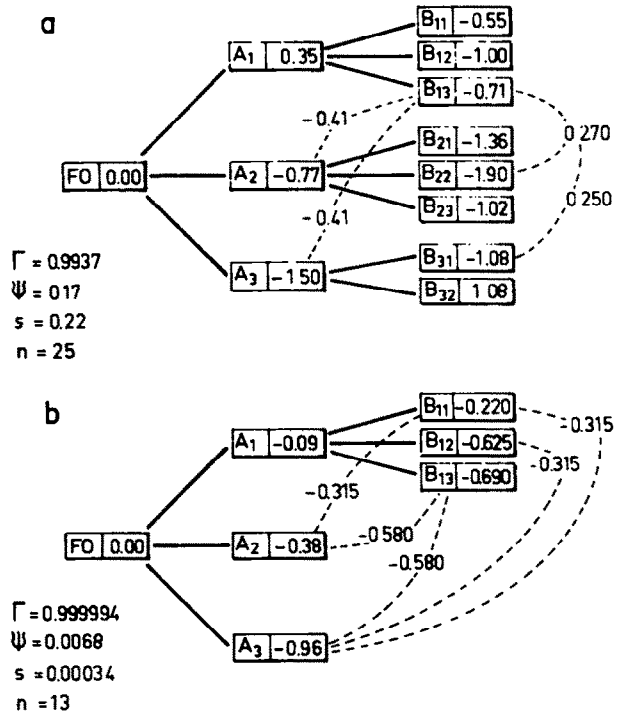
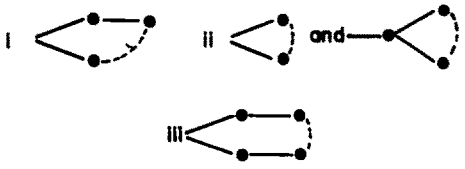


Fig. 5. Data treatment with site perturbations. Topological correlations for various populations. In this figure the perturbation of the overall property due to the introduction of a given site is indicated on the site itself whereas, in Fig. 1, the variation is indicated by arrows between particular structures. (a) Valuated PELCO graph for complete experimental population (correlation 1); (b) Similar valuated graph for population restricted to compounds in the trace of t-BuMe$_2$C- (correlation 2). The relevant E_S' values are given in Table 1.

actions. Henceforth, the term interaction will designate both implicit and explicit interactions, that is, interactions between sites that are dependent or independent, respectively, according to the DARC generation law. Thus the perturbations due to the introduction of the site A_2, $P(A_2)$, may be expressed as:

Perturbation $P(A_2)$ = Contribution $C(A_2)$
+ Interaction $I(A_2 - A_1)$

Here, we distinguish three types of interactions[22] between (i) sites of rank n and sites of rank n − 1; (ii) geminal sites of rank n; and (iii) non-geminal sites of rank n.



We also assume that the contributions of sites of the same rank have the same value, i.e.

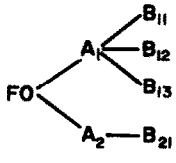
$$C(A_i) = C(A_r) \text{ and } C(B_{ij}) = C(B_{ij'})$$

Analysis of the E_S' values shows that interactions of the same type may be equated, except for the interactions $I(A_2 - B_{13})$ and $I(A_3 - B_{13})$ which differ from the other $I(A_i - B_{rj})$ interactions, i.e.

$$I(A_i - A_r) = I(A_i - A_r) \text{ and } I(B_{ij} - B_{ij}) = I(B_{rj} - B_{rr})$$

This gives rise to two site correlations, the first of which takes into account only primary and secondary groups (Corr. 3) while the second includes tertiary groups as well (Corr. 4). This treatment is developed according to the PELCO method with *a priori* interaction terms. The various contributions and interactions calculated by regression are given in Tables 2 and 3, and schematized in Fig. 6. The Student's test[24] on the regression coefficients for correlation 3 shows that, in spite of the fact that the overall correlation is significant, the data are not sufficient to determine the appropriate contributions and interactions. For correlation 4 the coefficients are highly significant, however, except for the slight positive contribution for the rank A carbons. This contribution does not differ significantly from zero. The good agreement between the known E_S' values not used in the correlation and the corresponding experimental values indicates that the unknown values calculated by this method are reliable. In correlation 4 the range of values is −0.48 to −5.29; the complete domain is 0 to −5.40.

As an example we may estimate the E_S' value of the group t-BuEtCH- using the data from correlation 4 (Table 3). The topological representation of this group is as follows:



2 C(A) + 1 I(A$_i$ − A$_r$) + 4 C(B$_{ij}$) +

3 I(B$_{ij}$ − B$_{ij'}$) + 3 I(A$_i$ − B$_{rj}$) +

1 I(A$_2$ − B$_{13}$) + 3 I(B$_{ij}$ − B$_{rj}$)

E_S' = −5.06 (Table 2)

Table 2. Topological correlation $\mathscr{E}/E'_S$ of alkyl groups—site contribution model with *a priori* interactions

GROUP	DEL	N° of C	E'_S exp	CORRELATION 3		CORRELATION 4	
				E'_S cal	Δ	E'_S cal	Δ
CH_3-	(1*)(0000)	1	0.00	0.00	0.00	0.05	0.05
$MeCH_2$-	(1*)(1000)	2	-0.08	-0.06	0.02	-0.02	0.06
$EtCH_2$-	(1*)(1100)	3	-0.31	-0.35	0.04	-0.28	0.03
Me_2CH-	(1*)(2000)		-0.48	-0.48	0.00	-0.48[a]	0.00
i-$PrCH_2$-	(1*)(1110)	4	-0.93	-0.88	0.05	-0.82	0.11
EtMeCH-	(1*)(2100)		-1.00	-1.00	0.00	-1.06	0.06
Me_3C-	(1*)(3000)		-1.43	-	-	-1.45	0.02
t-$BuCH_2$-	(1*)(1111)	5	-1.63	-1.64	0.01	-1.65	0.02
i-PrMeCH-	(1*)(2110)			-1.75	-	-1.92	-
Et_2CH-	(1*)(2200)		-2.00	-2.01[a]	0.01	-2.04	0.04
$EtMe_2C$-	(1*)(3100)		-2.28	-	-	-2.34[a]	0.06
t-BuMeCH-	(2*)(2111)	6	-3.21	-3.21	0.00	-3.27	0.06
i-PrEtCH-	(1*)(2210)		-3.23	-3.26[a]	0.03	-3.30	0.07
i-$PrMe_2C$-	(1*)(3110)		-3.54	-	-	-3.51	0.03
Et_2MeC-	(1*)(3200)			-	-	-3.63	-
t-BuEtCH-	(1*)(2211)	7		-5.21	-	-5.06	-
i-Pr_2CH-	(1*)(2220)		-5.01	-5.01	0.00	-4.97	0.04
Et_3C-	(2*)(3300)		-5.29	-	-	-5.33[a]	0.04
i-PrEtMeC-	(1*)(3210)			-	-	-5.21	-
t-$BuMe_2C$-	(1*)(3111)		-5.40	-	-	-5.38	0.02
i-$PrEt_2C$-	(1*)(3310)	8	-6.20			-7.32[b]	1.12
t-BuEtMeC-	(1*)(3211)					-7.49[b]	-
t-BuPr^{i}CH-	(1*)(2221)		-6.53			-7.13[b]	0.60
i-Pr_2MeC-	(1*)(3220)		-7.38			-7.20[b]	0.18
t-Bu_2CH-	(1*)(2222)	9	-6.97			-9.70[b]	2.73
t-$BuEt_2C$-	(1*)(3311)		-7.21			-10.00[b]	2.79
i-Pr_2EtC-	(1*)(3320)		-7.38			-9.71[b]	2.33
t-BuPr^{i}MeC-	(1*)(3221)		-7.56			-9.88[b]	2.32
t-BuPr^{i}EtC-	(1*)(3321)	10	-6.62			-12.79[b]	6.17
i-Pr_3C-	(1*)(3330)		-6.73			-12.50[b]	5.77
t-Bu_2MeC-	(1*)(3222)					-14.18[b]	-

a. The E'_S of this group not used in the correlation.
b. Values lying outside the range correlated.

Correlation 3 covers primary and secondary groups of Region I. Correlation 4 covers primary, secondary, and tertiary groups of Region I. It has a high predictive capacity in this region as shown by agreement between calculated values of those groups not correlated. Extension to groups of Regions II and III is presented for comparison purposes (Fig. 6).

Correlation 3: $r = 0.99989$, $\psi = 0.044$, $s = 0.026$, $n = 9$
Correlation 4: $r = 0.9995$, $\psi = 0.048$, $s = 0.057$, $n = 13$

where the appropriate interactions and contributions are given. This yields an E'_S value of -5.06 (Table 2).

The E'_S values calculated from correlation 4 for groups with more than 7 carbons are greater than the experimental values and the difference increases with the number of carbons (Table 2, Fig. 7). Only in the case of the group i-Pr_2Me– is the E'_S correctly estimated. It should be noted that the difference between the E'_S for this group and the group i-Pr_2CH–, from which it is derived by a formal replacement of H by Me is greater than the difference between the radicals with 6 and 7 C atoms. The satisfactory estimation of the E'_S of i-Pr_2MeC– is therefore not unusual.

(b) *$\mathscr{E}/E'_S$ correlation for groups with halogen in rank A.* This correlation method, based on the separation of the perturbation terms into contributions and interaction terms, is also applicable to restricted populations of groups halogenated in the α-position (rank A of the first ELCO). In order to have sufficient data, three isotopological populations containing Me and halogen (Cl, Br, or I) have been used. In this approach we separate the contribution of the reference carbon from the "chromatic" contribution of halogen. The results thus obtained are satisfactory:

$r = 0.996$, $\psi = 0.125$ (Cl groups)
$r = 0.996$, $\psi = 0.108$ (Br groups)
$r = 0.993$, $\psi = 0.170$ (I groups).

The values of the various contributions and interactions are given in Table 4.

Here again a slight positive contribution for the carbons of rank A, which statistically is not distinguishable from zero, is found. The reduced size of the population limits the predictivity (3 for the bromo, 1 for the chloro, and 3 for the iodo groups). The fluorine containing

Table 3. Contribution and interaction values of correlations 3 and 4

	Correlation 3	Correlation 4
Intercept	0.00	-0.055
C (A_i)	-0.065[j]	+0.035[j]
I ($A_i - A'_i$)	-0.350[h]	-0.500[b]
C (B_{ij})	-0.290[g]	-0.260[d]
i ($B_{ij} - B_{ij'}$)	-0.235[g]	-0.285[b]
I ($A_i - B_{i'j}$)$_{i'j\neq 13}$	-0.230[h]	-0.315[a]
I ($A_i - B_{13}$)	-0.695[g]	-0.520[a]
I ($B_{ij} - B_{i'j'}$)	-0.495[g]	-0.405[a]

confidence level of parameters :
(a) : > 99,9 % ; (b) : > 99,5 % ; (c) : > 99 % ; (d) : > 97,5 % ;
(e) : > 95 % ; (f) : > 90 % ; (g) : > 80 % ; (h) : > 60 % ;
(i) : > 50 % ; (j) : < 50 %.

See table 2

groups are not sufficiently numerous to merit such a treatment.

(c) *$\mathscr{C}$/E'_S estimation for alkyl groups extending to the second environment* (E_B^2). Table 5 gives the E'_S values for 14 alkyl groups extending into the second environment (E_B^2) of which 8 are included in the trace of $(CH_3CH_2CH_2CH_2)_2CH$–. These data are not sufficiently numerous to permit automatic calculation. However, the excellent results obtained for the first environment alkyl groups as well as the halogenated groups would seem to justify a non-automatic extension of the method to these groups.

As a first approximation we estimate the following contributions and interactions (Fig. 8):

$$C(A_i^{11}) = C(A_i^{21}) = C(B_{ij}^{11}) = C(B_{ij}^{21}) = -0.02$$

$$I(A_2 - A_i^{11}) = I(A_1 - A_i^{21}) = -0.21 \quad i = 2,3$$

In the second environment (E_B^2) the contributions of the sites of rank A are the same as those of the sites of

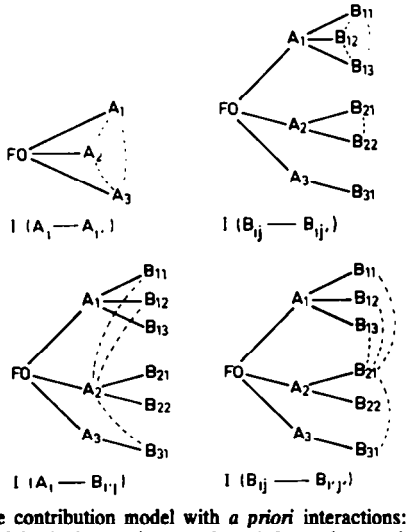


Fig. 6. Site contribution model with *a priori* interactions: enumeration of basic interactions evaluated from the topological correlations. For the two upper graphs all relevant interactions ($I(A_i - A_{i'})$ and $I(B_{ij} - B_{ij'})$) are shown. For the two lower graphs only a few of the possible interactions ($I(A_i - B_{i'j})$ and $I(B_{ij} - B_{i'j'})$) are indicated, in the interest of clarity. All interactions of the same general type are equivalent except in the case of the interactions $I(A_i - B_{ij})$ where $I(A_2 - B_{13})$ are equivalent to each other but different from the others of this type.

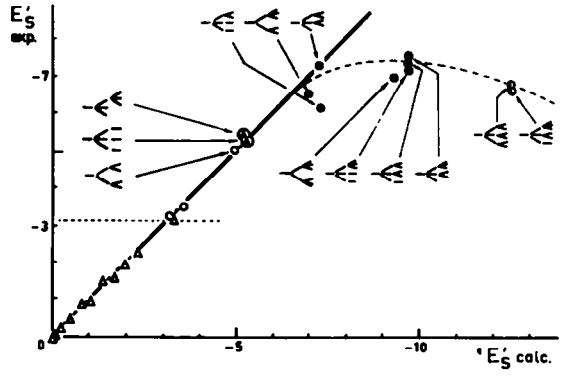


Fig. 7. Relationship between experimental E'_S values and those calculated from site contribution with systematic interaction model (correlation 4). The *augmentation* (linear portion), *levelling* and *inversion* effects are evident. △—values derived from literature data. ○, ●, ◐, new E'_S values corresponding to regions I, II and III of Fig. 1, respectively. ⓐ—values notably in conflict with previous estimates.

Table 4. Topological correlation—site contribution model with *a priori* interactions as applied to groups with halogen atoms in rank A

	X = Cl			X = Br			X = I		
Group	E'_S exp	E'_S cal	Δ	E'_S exp	E'_S cal	Δ	E'_S exp	E'_S cal	Δ
CH_3-	0.00	0.04	0.04	0.00	−0.04	0.04	0.00	−0.04	0.04
$MeCH_2$-	−0.08	−0.03	0.05	−0.08	0.00	0.08	−0.08	+0.04	0.12
X CH_2-	−0.18	−0.12	0.06	−0.24	−0.20	0.04	−0.30	−0.29	0.01
Me_2CH-	−0.48	−0.50	0.02	−0.48	−0.49	0.01	−0.48	−0.44	0.04
Me X CH-		−0.59			−0.69			−0.77	
X_2CH-	−0.58	−0.68	0.10	−0.76	−0.89	0.13	−0.93	−1.11	0.18
Me_3C-	−1.43	−1.45	0.02	−1.43	−1.51	0.08	−1.43	−1.50	0.07
Me_2XC-		−1.54		−1.77	−1.71	0.06		−1.83	
MeX_2C-		−1.63		−1.92	−1.91	0.01		−2.17	
X_3C-	−1.75	−1.72	0.03	−2.14	−2.11	0.03	−2.62	−2.50	0.12
	C_C = + 0.01 [j]		r = 0.9966	C_C = + 0.04 [j]		r = 0.9967	C_C = + 0.09 [j]		r = 0.9933
	C_{Cl} = − 0.09 [f]		ψ = 0.125	C_{Br} = − 0.20 [b]		ψ = 0.108	C_I = − 0.335 [c]		ψ = 0.17
	I = − 0.48 [b]		s = 0.061	I = − 0.53 [a]		s = 0.073	I = − 0.575 [d]		s = 0.118

(a), (b).. . (j) cf. Table 3 .

Because of the small amount of data available for halogenated groups we have used a mixed data set (methylated and halogenated).

Table 5. Topological estimation—site contribution model with *a priori* interactions for alkyl groups with four concentric layers of carbon atoms

Group	DEL	E_S' exp	E_S' cal	Δ
n-$PrCH_2$-	(1*)(1100*)(1000)11	-0.35	-0.33	0.02
n-$BuCH_2$-	(1*)(1100*)(1100)11	-0.31	-0.35	0.04
n-PrMeCH-	(1*)(2100*)(1000)11	-1.02	-1.02	0.00
n-BuMeCH-	(1*)(2100*)(1100)11	-1.06	-1.04	0.02
n-PrEtCH-	(1*)(2200*)(1000)11	-2.00	-2.02	0.02
n-BuEtCH-	(1*)(2200*)(1100)11	-2.03	-2.04	0.01
$(n\text{-}Pr)_2CH$-	(1*)(2200**)(1000)11,21	-2.03	-2.04	0.01
n-$BuPr^nCH$-	(1*)(2200**)(1100)11(1000)21		-2.06	
$(n\text{-}Bu)_2CH$-	(1*)(2200**)(1100)11,21	-2.08	-2.08	0.00
i-$BuCH_2$-	(1*)(1100*)(2000)11	-0.33	-0.35	0.02
t^a-$BuCH_2CH_2$-	(1*)(1100*)(3000)11	-0.32	-0.37	0.05
i-BuMeCH-	(1*)(2100*)(2000)11		-1.25	
t-$BuCH_2$MeCH-	(1*)(2100*)(3000)11	-1.81	-1.48	0.33
i-BuEtCH-	(1*)(2200*)(2000)11		-2.25	
t-$BuCH_2$EtCH-	(1*)(2200*)(3000)11		-2.48	
i-$BuPr^nCH$-	(1*)(2200**)(2000)11(1000)21		-2.27	
t-$BuCH_2Pr^nCH$-	(1*)(2200**)(3000)11(1100)21		-2.50	
$(i\text{-}Bu)_2CH$-	(1*)(2200**)(2000)11,21	-2.38	-2.52	0.12
t-$BuCH_2Bu^iCH$-	(1*)(2200**)(3000)11(2000)21		-2.73	
$(t\text{-}BuCH_2)_2CH$-	(1*)(2200**)(3000)11,21	-3.06	-2.96	0.10
n-$PrMe_2C$-	(1*)(3100*)(1000)11		-2.30	
i-$BuMe_2C$-	(1*)(3100*)(2000)11		-2.53	
t-$BuCH_2Me_2C$-	(1*)(3100*)(3000)11	-2.48	-2.76	0.28

The site values of E_B^2 are estimated since the data are insufficiently numerous for a systematic correlation.

rank B. In the case of primary and secondary linear groups no interactions are considered for these sites due to their remoteness from the reaction centre. The maximum difference between calculation and experiment is 0.33 for the group CH_3(t-$BuCH_2$)CH–, representing an error of 18%; the average difference is 0.07 while the range of experimental values extends from −0.35 to −3.06. This approach allows a reasonable estimate of 9 new E_S' values.

The topological correlations $\mathscr{C}/E_S'$ which relate individual sites and overall group parameters allow the prediction of new values and bring out the relative importance of site contributions and interactions between non-bonded atoms. To go further in our understanding of the mechanism by which the steric effect operates the topology, which represents the topography only in the case of relatively simple molecules, must be completed by a study of conformation. In particular, the fact that the interactions $I(A_2 - B_{13})$ and $I(A_3 - B_{13})$ are numerically larger than the other interactions $I(A_i - B_{i'j})$ suggests that changes in preferred conformations are operating. This is quite a detailed problem in itself and will be considered in the following article.

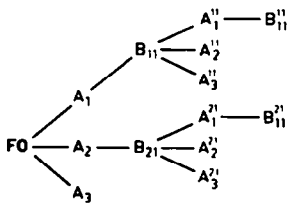


Fig. 8. Trace of experimental set of alkyl groups belonging to the second limited environment.

Overview

The generative idea in this article is the concept of topological sequencing which is used in a very specific way to correlate and predict the evolution of steric effects of alkyl groups with substitution.

The PELCO topological approach, consisting of successive refinement stages, begins with a very general model (Correlation 1) which is able to correlate data over a range of 7 log units (Regions I, II and III) but has only slight predictive capacity. Restricting our attention to the data of Region I and choosing a simplified *a priori* model yields a correlation which can be used to predict reliably steric effects within this limit (Correlation 4). This approach has been adapted to encompass groups containing

α-halogen atoms and those extending into the second environment (E_B^2).

In the following article the ideal of sequencing will be pursued further, this time at the conformational rather than at the topological level, making use of, among other things, the empirical force field method. While the accent in this article has been on *correlation* and *prediction*, in the following one it will be on the closely related problem of *interpretation* in mechanistic terms.

REFERENCES

[1a]Maître-Assistant, Université Paris VII; [b]Chargé de Recherche, C.N.R.S.

[2]J. A. MacPhee, A. Panaye and J. E. Dubois, *Tetrahedron* **34**, 3553 (1978); *Tetrahedron Letters*, 3293 (1978).

[3a]Description, Acquisition, Retrieval, Computer Aided Design; [b]J. E. Dubois, D. Laurent and H. Viellard, *C. R. Acad. Sci.* Ser. C, **263**, 764 (1966); [c]J. E. Dubois, *Entropie* **25**, 12 (1969); [d]J. E. Dubois and H. Viellard, *Bull. Soc. Chim. Fr.* 900, 905 and 913 (1968); [e]J. E. Dubois, H. Viellard and A. Panaye, *Ibid.* 1988 and 1996 (1973); [f]J. E. Dubois and A. Panaye, *Ibid.* 1390, 1401 and 2100 (1975); *Ibid.* 1229 (1976).

[4a]J. E. Dubois, D. Laurent and H. Viellard, *C.R.Acad.Sci.* Ser.C, **264**, 1019 (1967); [b]J. E. Dubois, D. Laurent and A. Aranda, *J. Chim. Phys.* 1608 and 1616 (1973).

[5]J. E. Dubois, J. P. Anselmini and M. Chastrette, *Bull. Soc. Chim. Fr.* 2439 (1969).

[6a]J. E. Dubois, D. Laurent, A. Panaye and Y. Sobel, *C. R. Acad. Sci.* Ser. C, **281**, 687 (1975); [b]J. E. Dubois, D. Laurent, A. Panaye and Y. Sobel, *Ibid.* Ser. C, **280**, 851 (1975).

[7a]J. E. Dubois and A. Panaye, *Tetrahedron Letters* 1501 (1969); [b]*Ibid.* 3275 (1969).

[8]A. T. Balaban, D. Farcasiu and R. Banica, *Rev. Roum. Chim.* **11**, 125 (1966).

[9]C. K. Johnson, C. J. Collins, *J. Am. Chem. Soc.* **96**, 2514 (1974).

[10]J. B. Hendrickson, *Ibid.* **89**, 7047 (1967).

[11]L. M. Masinter, N. S. Sridharam and D. H. Smith, *Ibid.*, **96**, 7702 (1974).

[12]J. E. Leffler and E. Grunwald, *Rates and Equilibria of Organic Reactions*. Wiley, New York (1963).

[13a]J. E. Dubois and H. Herzog, *Chem. Commun.* 932 (1972); [b]J. Chrétien and J. E. Dubois, *J. Chromatogr.* **126**, 171 (1976); [c]J. Chrétien and J. E. Dubois, *Analyt. Chem.* **49**, 747 (1977).

[14a]A. Massat and J. E. Dubois, *J. Mol. Struct.* **4**, 385 (1969); [b]J. E. Dubois, A. Massat and P. Guillaume, *Ibid.* **4**, 403 (1969); [c]J. E. Dubois, J. P. Doucet and B. Tiffon, *J. Chim. Phys.* 805 (1973); [d]J. E. Dubois and J. P. Doucet, *J. Org. Magn. Res.*, in press.

[15a]J. E. Dubois, *Chim. Ther.* 1, 65 (1972); [b]A. Aranda, *C. R. Acad. Sci.* Ser. C, **276**, 1301 (1973); [c]J. E. Dubois, *Man and Computer* (Edited by M. Marois), p. 309. North Holland, Amsterdam (1974); [d]J. E. Dubois, D. Laurent, P. Bost, S. Chambaud and C. Mercier, *Eur. J. Chem., Chim. Ther.* **11**, 225 (1976).

[16]C. Bergé, *Théorie des graphes et ses applications*. Dunod, Paris (1963).

[17a]ELCO: *Environment which is Limited to two layers of atoms Concentric around a focus, Ordered by a given set of generative rules.* [b]J. E. Dubois, *The Chemical Application of Graph Theory*, (Edited by A. T. Balaban), p. 336. Academic Press, London (1976).

[18]D. H. Smith and P. C. Lurs, *J. Am. Chem. Soc.* **100**, 3316 (1978).

[19a]J. E. Dubois, M. J. Alliot and H. Viellard, *C. R. Acad. Sci.* Ser. C, **271**, 1412 (1970); [b]J. E. Dubois, M. J. Alliot and A. Panaye, *Ibid.* Ser. C, **273**, 224 (1971); [c]J. E. Dubois and M. J. Cojan-Alliot, *Ibid.* Ser. C, **280**, 13 (1975); [d]J. E. Dubois, A. Panaye and M. J. Cojan-Alliot, *Ibid.* Ser. C, **280**, 353 (1975).

[20]J. E. Dubois and A. Panaye, *Bull. Soc. Chim. Fr.* 1229 (1976).

[21]PELCO: Perturbation of an ELCO, see Ref. 4*b*.

[22]J. E. Dubois, A. Panaye and J. MacPhee, *C. R. Acad. Sci.* Ser. C, **280**, 411 (1975).

[23]J. E. Dubois, J. A. MacPhee and A. Panaye, *Tetrahedron Letters* 4099 (1978).

[24]Statistical tests used in this article: S—standard error of estimate; r—correlation coefficient; ψ—Exner's test, I. Exner, *Coll. Czech. Chem. Commun.* **31**, 3222 (1966); for F test and student t see, G. W. Snedecor *Statistical Methods* (5th Ed.) the Iowa State College Press, Ames, Iowa (1956).