Pure & Appl. Chem., Vol.53, pp.1313-1327.
Printed in Great Britain.

0033-4545/81/061313-15$02.00/0
Pergamon Press Ltd

COMPUTER ASSISTED MODELLING OF REACTIONS AND REACTIVITY

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

Abstract - With the surge of computers, the representation of chemical changes needs a broader space than that provided by traditional approaches. Reaction modelling and reactivity modelling are two arbitrarily distinguished aspects for which, hopefully, the treatments will converge in the future. For reactions, it is shown how the DARC topological system makes possible a faithful written description of reactions through the notion of generative grammar underlying the DARC system. Structural changes in a reaction are described according to the DARC-IGLOO method, whereas the scope of a reaction is described and quantified by means of the Domain of a Structurally Ordered Reaction, DSOR. For reactivity, the restriction imposed by the limited representation spaces and faulty parametrization of classical single or multi-parameter LFER are surmounted by the introduction of interaction functions into these LFER so that non-additive phenomena can be considered, and by the DARC-PELCO topological correlation method based on structural site parameter estimates for reactions embedded in ordered hyperstructures (formal graphs).
Progress in representation and reactivity strategies should allow a better local and global comprehension for future reactivity correlations.

INTRODUCTION

Chemistry is a science that deals both with substances and the changes they undergo. The reactivity of a chemical compound, i.e. its ability to undergo change in the course of a reaction, is actually a complex function of its properties. However, studies on the structure and on the taxonomy of chemical compounds are not only greater in number and date back much earlier than those dealing with chemical changes, but they have also recently benefitted to a greater extent from physical methods.

For a good while the study of reactions was restricted to that of initial and final products. Then, the stages of change were approached from the standpoint of general mechanisms by gathering kinetic data, or less formally from the standpoint of specific mechanisms. Early models also benefitted from physical and structural studies of intermediate species. The space in which chemical changes are represented was thus partly defined by the structure of the reactants, products and intermediate species. With the sudden surge of computers, the broadening of this representation space has seemed desirable in order to shed further light on the problems of chemical change.

It should be remembered that whatever concepts of reaction and reactivity are to be used in modelling, they must first be placed in a well-defined, flexible and coherent representation space. This requirement always entails more rigorous comprehension of the concepts, phenomena and their taxonomies. It will probably give rise to axioms suited to data related to the dynamics of chemical systems.

The aptitude for modelling data to be handled is closely related to the extent of their generality (i.e. the broader the data, the harder it is to handle them accurately) and to the compromises accepted in taking them into account. Thus, in certain attempts at computer-aided synthesis (CAS), data banks have been built with simplified or usual descriptions of structural or reaction data. The software using these data makes it possible to represent certain models which define a simulation space that is mutilated to start with by the choice of representation. In every instance of modelling, a clear distinction must be made between *the representation space of the data* and the set of manipulations which can be achieved by the various treatments of these data, often considered as a *treatment space*. The strategies of computer-aided design (CAD) can be limited by the strength of the treatment operations, but this can be only temporary. In contrast, the loss of information when designing the *representation space* is irretrievable.

In discussing reaction modelling, an attempt is made here to dissociate in a rather arbitrary manner the concept of reaction from that of reactivity (Fig. 1), and to show that, in each case, the current representation or treatment frameworks impose constraints which limit the representation space, the essential space of modelling.

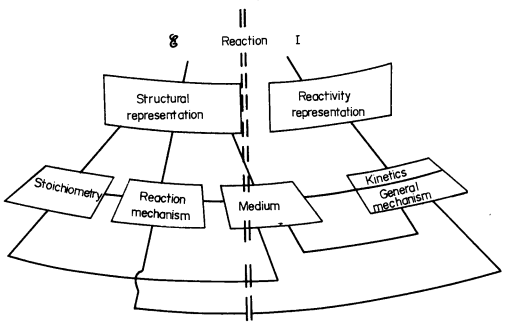


Fig. 1. Relationship Between Reaction and Reactivity.
The diagram shows the overlap between characteristics associated with the reaction concept and those associated with the reactivity concept. Representation in graph form rather than tree form illustrates the complexity involved in the problem of treating these concepts separately.

For reactions, my intent is to show how the DARC topological system makes possible a faithful written description of reactions, through the notion of generative grammar which underlies the DARC system. For reactivity, which associates the system of changing structures with solvent parameters, I question the over-simplicity of standard kinetic treatments.

Given the complexity of simulations and the wish to unify reaction treatments without subjecting them to outrageous simplification, the efforts called for to build a vaster representation space and treatment space are justified by the hope or need to rely on computer assistance.

REACTION MODELLING

The modelling of reactions presented herein fits in the general framework of the DARC system (1) in which an attempt has been made to elaborate a coherent language for structural data related to the various chemical substances, compound populations and reactions; it also prepares for the various activities of chemical informatics (2).

Computer-aided design (CAD) problems in chemistry can be approached via three main axes (3): information science for logistics and documentation, logic related to structural concepts and combinatorics, and language (in the present case, the nomenclature and codes when lend potential to the data). To begin with, users of CAD have brought their efforts to bear on information science by transferring the usual methods of thought employed by chemists to computer-perceptible programs. However, it seems that a wider opening should first be made for the logic and a certain language in chemistry, in order to benefit from the intuition of chemists. This opening is aimed at improving the heuristic potential sought in computer-aided synthesis (CAS) as well as in computer-aided correlation (CAC) and structure elucidation (CAE). Over the last ten years, these three areas of chemistry have been studied in an isolated manner by various research teams (4).

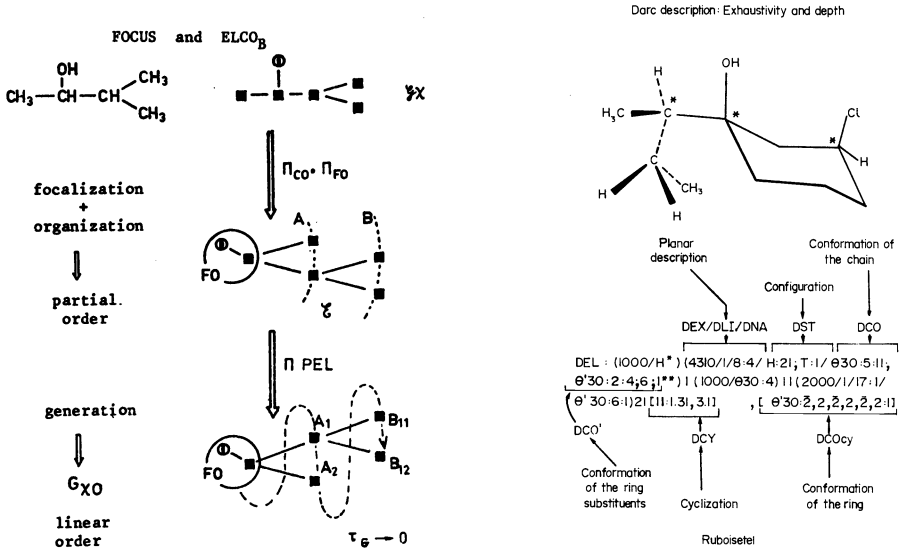
However, it should be noted that data banks and the various areas of computer-aided design require common general tools as well as related finalized tools. For this reason it seems that generality and adaptability should be kept in mind when elaborating such tools and that this is a requirement for surmounting the difficulties encountered in chemical informatics. Generality is needed in order to benefit from the potentials and development acquired in a given field; adaptability is needed because one must not forget that the representation space adopted for the construction of a data bank is related to its use.

DARC Treatment of Compounds

Each structure S is formally assimilated to a chromatic graph whose nodes represent non-hydrogen atoms and whose edges represent the bonds between these atoms. In the graph, two parts can be distinguished: the focus (FO), made up of an atom, a bond, or a group of atoms

characterizing a series of compounds, and the environment (&) which brings together the remaining sites (atoms, bonds) that are concentrically organized around the focus (5,6).

Generation principle. Grasping a structure by means of generation consists in creating the structure representation through an ordered process which reflects the sequential steps involved in building the structure by means of two elementary operations: ablation (ab) or adjunction (ad) of topological or chromatic information. Generation occurs by selecting an origin or focus (FO) and building the environment (&) through gradual and ordered substitutionsof the hydrogen atoms (Fig. 2).



(a) (b)

Fig. 2. Generation pathway and total ordering on structure. (A chemical compound is modeled by a chromatic graph G_χ whose nodes represent the heavy atoms (carbon + heteroatoms) and whose edges represent the bonds between these atoms . (a) The graph is organized concentrically (π_{OC}) around a point or a set of points called Focus (π_{FO}). The environment &, or set of sites outside the focus is ordered linearly by applying topological and chromatic criteria. (b) Then the focus and its environment are expressed by a linear code, the DEL.

The generation automata creates, starting from a focus (FO), the graph $G_{\chi O}$, by taking the graph G_χ as its target. A topochromaticity table and a set of rules allow priority decisions, and generation thus operates according to a Major Order (M.O.).

All the graphs created during the generation of the chromatic target graph constitute a defined family whose members are well located in relation to each other (Fig. 3). The whole

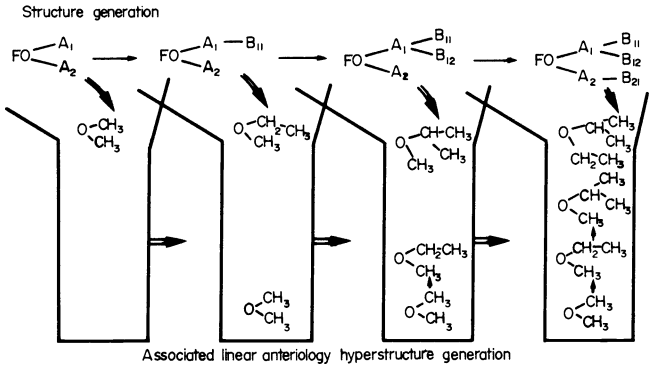


Fig. 3. Relationship Between the Progressive Construction of a Structure (S) and That of Its Associated Hyperstructure (HS) in which Structural Filiation Relations are Memorized.
Each time a new structure is generated, it joins HS which grows synchronously with S.

of their respective relationships forms a network or graph in which the family members are the nodes. In the DARC system there is a close synchronous relationship between the progressive construction of the structure (S) and that of the associated hyperstructure (HS) (7). Indeed, while a structure is being generated, the generation is expressed by the ordered adjunction (ad_o) of an edge and a site into a structure (S). One thus obtains a structure called S'. In this way, for each ordered adjunction of an edge and a site into a structure, there is an ordered adjunction of an edge and a site into a hyperstructure (HS). In applying the principle of synchronism, procedures for enumerating and listing isomers and related compounds by anteriority relationships (7) have been elaborated. (Fig. 4)

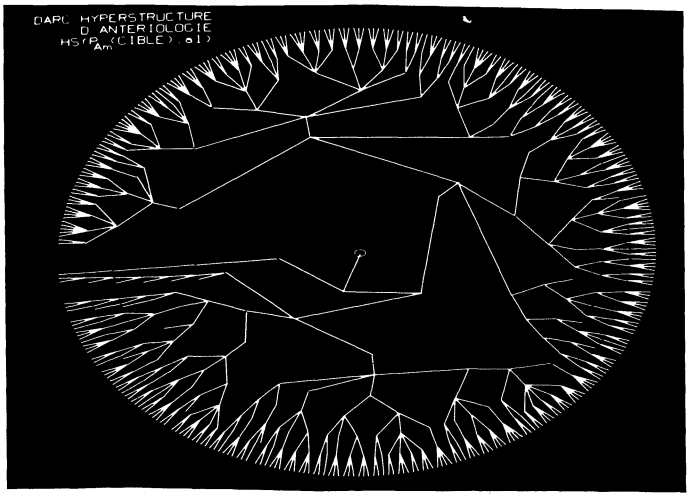


Fig. 4. Anteriology Hyperstructure for C_nH_{2n+2} Alkanes (n = 1-12).
By means of the ordered adjunction of sites, the generation of the isomers at each level of isomerism is achieved without redundance. This generation simultaneously solves the problem of enumerating and listing the set of structures in orderly fashion (664 structures, 355 of which have 12 atoms).

DARC Treatment of Reactions

The representation of reactions discussed here is that used in the DARC system to deal with problems related to computer-aided synthesis (CAS) and computer-aided elucidation (CAE). A different approach is used in specific reaction banks. In either case, data concerning the compounds involved in a reaction must be separated from data concerning the reaction medium. The topic which will be dealt with here is structural data concerning compounds involved in a reaction. However, it should be briefly noted in passing that solvent effect data in certain DARC data banks are grouped together in a vector in which each component is set at 1 or 0, depending on whether or not there is a corresponding characteristic.

In accordance with the DARC rules for describing structures, the DARC-IGLOO (8,2c) method makes it possible to associate a biunivocal representation with all the reactions in organic chemistry (Fig. 5). The search for a unique order for a reaction has led to characterizing a reaction on the basis of its modelling by a set of initial and final graphs. Superposing the graphs of the initial and final compounds yields the Greatest Common Sub-Structure (GCSS) (9) from which it is possible to isolate the pivot focus (FOP), i.e. the smallest connected set of nodes serving as ties for structural variations. The initial and final compounds are respectively reconstituted by ADjunction Operators (ADO) and ABlation Operators (ABO).

Since 1978, the possibility to describe unconnected graphs associated with the reaction invariant has allowed reactions of the type m:n (i.e. containing any number of initial and final products) to be handled.

Available information can be handled in a domain ranging from the highly specific (description of a well-defined reaction which is in the literature) to the highly generic (the broadest overall description possible, eliciting a study of the range of applicability of a general reaction such as the Mannich reaction).

The notion of the unique order of a reaction was used in building the software (PARIS program) (10) which, depending on the interactively chosen options, makes it possible to generate either syntheses or antitheses. This original approach permits introducing the concept of a concurrent reaction which takes a priori structural criteria into account and leaves the user the choice of constraints related to the chemical reality.

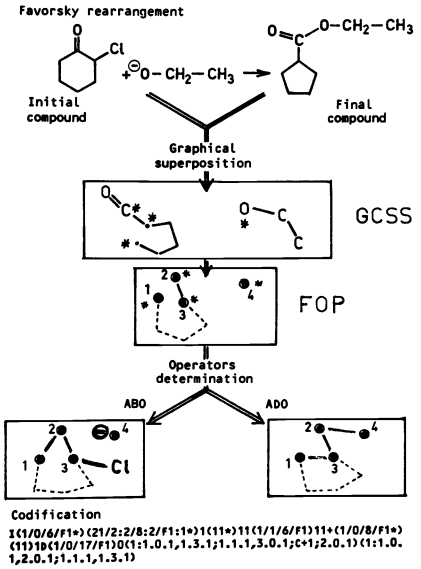


Fig. 5. Specific Reaction by IGLOO Method. By the IGLOO method (Invariant Graph and Localized Ordered Operators), change is isolated with reference to the Greatest Common Sub-Structure invariant of the reaction (GCSS). From this GCSS, we isolate the pivot focus which groups the atoms of the GCSS, sites of adjunction or ablation. Once the pivot focus is ordered, the Invariant and Localized Operators (ILO) (Operators of ablation ABO and of adjunction ADO) and the INvariant which is Concentric and Ordered (INCO) make it possible to establish the DILO Descriptor of the Reaction .

In order to assess the potential of a generic reaction on the basis of relevant specific reactions, without having to rely on an a priori interpretation, the following notions have been proposed: the Domain of a Structurally Ordered Population, DSOP, which includes those structures located between the lower and upper limits (i.e. between root compounds and summit compounds) of a population P of compounds related by formal algorithmic filiations; the Domain of a Structurally Ordered Reaction, DSOR, described as a function of the DSOP and of the reaction constraints; and, given certain external parameters, an extrapolation of the latter domain, $DSOR_E$. Several types of proference are used to express the potential of the methods of synthesis (11).

These notions have been illustrated by a study of two synthesis reactions of aliphatic ketones which are limited to position γ [the Grignard reaction for acid chlorides (D_{3Ai}) and alkylation (B_1)] . Using a 47-compound key population, the Extrapolated Domain of a Structurally Ordered Reaction, $DSOR_E$, for the Grignard reaction (D_{3Ai}) is assessed at 223 ketones, whereas using a 44-compound key population for ketone alkylation (B_1) it is assessed at 344. The set of known aliphatic ketones with at most some carbon atoms at position γ with respect to the carbonyl group determines a population of 416 ordered ketones limited by 9 summit ketones (7 of which are synthesized here). Aside from the 23 commercial ketones, the union of $DSOR_E$, D_{3Ai} and B_1 indicates that 381 ketones can be synthesized by either one of the two above-mentioned methods.

In the hyperstructure diagram for the Grignard reaction (Fig. 6), the 11 summit compounds and the 6 root compounds determine a population domain of 204 ketones. In view of the inherent constraints of the method and the possible extrapolations, the $DSOR_E$ of method D_{3Ai} contains 223 ketones. The 47 compounds of the key population are located at the boundaries of each partial domain: primary-secondary ketone, primary-tertiary,... . The ketones in the blank squares belong to DSOP, but are excluded because of the difficulty involved in obtaining secondary magnesium compounds. The ketones in the blank areas lie beyond the structural boundary determined by the summit ketones.

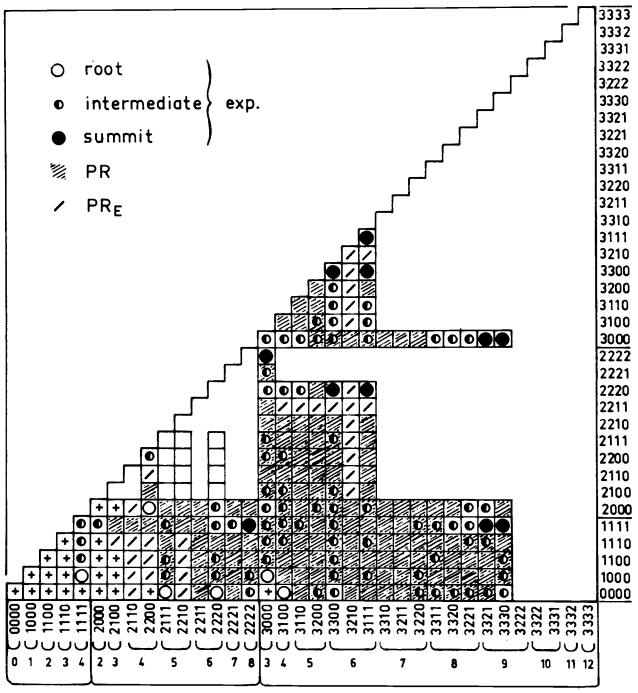


Fig. 6. Domain of a Structurally Ordered Reaction (DSOR) for the Grignard Reaction on an Acyl Chloride (D_{3Ai}).
Each square of a hyperstructure diagram corresponds to a compound. The code-numbered environment (DEL) of the two directions of development of the focus chosen for the ketones, $\not C\text{-}CO\text{-}C\not$, is indicated along the axes of the diagram. The finalized order used to classify the DEL is that best adapted for a closer analysis of the D_{3Ai} method.

REACTIVITY MODELLING

Modelling means structuring reactivity data logically at the acquisition and description level in order to facilitate its storage and retrieval without subjecting it to alteration or mutilation. On a less formal level, modelling is the basis for any prediction of, or any attempt at understanding, reactivity. The classification of the modelling methods presented here highlights the fact that understanding and prediction, either separately or jointly, have been the underlying goals of every method involving the handling of reactivity-related data.

The reactivity of a system is quantitatively assessed by the rate constant k_{ij} of the reaction which transforms the reactant X_j - R into product P.

$$X_j - R \xrightarrow[k_{ij}]{M_i} P$$

Reactant Product

By its very definition, the reactivity depends on the structure of the reactant and on the conditions of the reaction medium. Thus, reactivity is a function of two types of variables:

$$\log k_{ij} = F(X_j, M_i)$$

- internal variables which express the influence of the structure of the reactant (e.g. substituents, stereochemistry,...),
- external variables which describe the medium (e.g. solvent, temperature, catalyst, pressure,...).

Therefore, modelling involves (i) defining the representation space, (ii) parametrizing the variables, i.e. assigning numerical parameters to non-quantitative factors, and (iii) setting up the function F which relates these parameters to the reactivity.

The representation space is generally determined by the two vectors associated with the

structure and the medium, and each vector is a complex function of n variables (12). To determine function F a first approach involves separating out the effect of each variable by developing F classically:

$$\log k_{ij} = F(X_j, M_i) = \log k_\infty + f^1_{M_i}(X_j) + f^2_{X_j}(M_i) + \phi(X_j, M_i)$$

This generally leads to treating the effect of each variable (X_j or M_i) separately. It should immediately be noted, however, that knowing the effect of each variable is not enough to correctly describe the reactivity, since aside from function f_i there is an additional function ϕ describing the interactions between each pair of variables.

The most frequently used empirical process for describing the effect of a single variable upon the reactivity is the method of linear free energy relationships. Indeed, with ease of handling in mind, the reactivity is dealt with in its simplest form, that of linear or multilinear functions. Likewise, for automatic handling by computer, the polynomial functions used are linear models of the variables which can be complex n-order functions of the parameters describing them.

The parametrization of variables related to structure or medium (the other aspect of modelling) is a more complex matter and has been dealt with in several ways, depending on the objectives in mind. Chemists as well as information scientists tend to devote more attention to parametrization than to the choice of functions. The different types of modelling related to parametrization which have been or can be used, as well as the limits of applicability of each method, are discussed hereafter.

Modelling a Variable by a Single Parameter

The few models used in theoretical chemistry for simple molecules reacting in the gas phase are of little use to experimental chemists studying complex molecules reacting in solution. Many empirical methods have been developed to deal with the latter situation.

The simple models which have been implemented in experimental chemistry are most often highly specific and therefore reproduce the information well, although in a restricted context. A consequence of the specificity of these models is that a great many scales of parameters have been proposed for structure effects as well as for solvent effects. Although Tables 1 and 2 list only some of the scales proposed [the reader wishing for a more exhaustive list should consult the recent reviews by Exner (13,15) or Reichardt (18)], they do indicate the bases on which these models are built and which largely determine their field of application.

TABLE 1. Modelling of Structural Effects

GROUP/EFFECT [a]			
Parameters from	Reactivity data	: σ^*, E_s, ...	, Taft
		$\sigma, \sigma^+, \sigma^-$	, Hammett , Brown
	Spectroscopic data	: σ_I, σ_R°, ...	, Taft
	Physical data	: γ	, Charton
		B_1, B_2, B_3, B_4...	Verloop
		E_{FMO}	Fukui, Klopman
SITE/EFFECT [b]			
	Topological parameters	: DARC-PELCO	Dubois

a) Extensive reviews on these parameters are given in refs. 13, 15, 16 and 19.
b) Ref. 29.

Physical Models. The first models proposed for the parametrization of the variables were based on physical or physicochemical properties of a medium or of structural groups. Such is the case for the models proposed by Charton (19) (structural parameter ν related to the Van der Waals radius), Kirkwood (20) (solvent parameters related to the dielectric constant, or to the refractive index) and Parker (21) (solvent parameter ΔG_{tr} related to ion transfer energies).

It is obvious that these models based on a unique macroscopic physical parameter are too simple to account for the usual complexity of the effects on reactivity and therefore have only been applied in specific cases.

Analogical Models. In complete contrast to the above models, in analogical models the parameters are defined by a reference model without relying on a hypothesis concerning the nature of the factors affecting the reactivity. The highly diverse scales of parameters in this field have been widely applied. The reference processes for structural parameters (13,14,15,16) such as Hammett's σ, the Hammett-Brown σ^+ and Taft's σ^*, and for solvent effect parameters (18) such as Winstein's Y or N, are reactivity processes. In contrast, the reference processes for solvent parameters such as those based on shifts in λ_{max} (E_T, Z or ϕ scales) or for structure parameters based on NMR chemical shifts (δ) are spectroscopic processes (18) (Tables 1 and 2).

TABLE 2. Modelling of Solvent Effects

Parameters [c] from			
	Reactivity data :	Y, N, ...	Winstein, Grunwald
		d_1, d_2 ...	Swain, Lupton
		DN	Gutmann
	Spectroscopic data :	AN	Gutmann
		E_T, Z, ϕ,.	Dimroth, Kosower, Dubois
		α, β, π^*,.	Taft, Kamlet
		G.....	Schleyer, Allerhand
	Physical data :	ε, n, μ ..	Kirkwood
		ΔG_{tr}	Parker

c) Extensive reviews on the definition of these parameters are given in Refs. 14,18.

The extraordinary variety of parameters of this type and the diversity of their applications demonstrate both their success in expressing reactivity and their lack of general significance. Indeed, in applying these empirical parameters, it is implicitly assumed that the microscopic factors acting on the reactivity are the same in the reference process as in the reaction studied. Obviously, this is only valid for very similar processes. Consequently, it cannot be expected that an empirical scale based on a specific process will be universal.

Using Analogical Models (Example). When analogical models are used within the limits of their field of application their main advantage lies in that they highlight similarities between true reaction processes. For example, the structure of the intermediate in the bromination of α-methylstilbenes can be identified (22) by $\rho\sigma$ analysis of the reactivity. The effects of the ring-substituents on the reactivity are correctly modelled by the σ^+ constants derived from methanolysis of tert-cumyl chlorides. In both reactions, the rate-limiting step is the formation of a benzylic carbocation.

The following example, taken from the solvent effect on the bromination of 1-pentene (23), also shows that analogical models are easier to apply than physicochemical models.

The solvent effect function for a physical parameter such as Kirkwood's $\frac{\varepsilon - 1}{2\varepsilon - 1}$ function is very complex and, in any case, is not linear. One comes close to linearity when modelling with E_T. In contrast, a satisfactory linear function is obtained by using the parameter Y. It follows that in bromination as in the reaction defining Y (solvolysis of tert-butyl chloride) the main role of the solvent is leaving group assistance (24) (Fig.7).

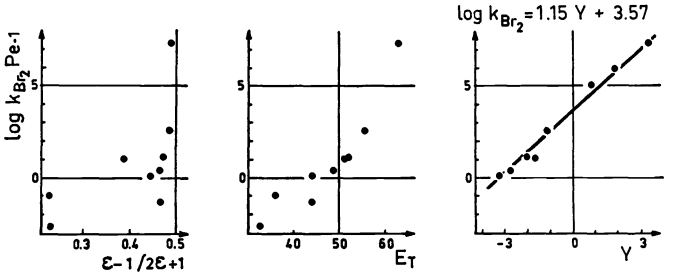


Fig. 7. Search for a Linear Function in the Modelling of Solvent Effects on the Reactivity of 1-Pentene with respect to Bromine. The solvents, in decreasing order of influence, are: H_2O, MeOH, EtOH, AcOH, iPrOH, t-BuOH, dioxane, Ac_2O, CCl_4.

Limits of Analogical Models. In all cases, simple linear models involving limited representation space can only represent limited populations, either because the number of parameters in a given scale is insufficient, or because, in a large population, more than one factor acts on the reactivity. Therefore, other more complex types of modelling must be sought.

Multiparameter Models of a Single Variable

Since the limits of analogical models partially result from the lack of hypotheses on the factors influencing reactivity, models based on the identification and separation of these factors have been sought. The overall effect on reactivity is assumed to be a linear combination of the functions of the parameters related to these microscopic factors. In this way, these approaches correspond to an extension of the representations space.

Such a modelling methodology involves several assumptions:
- that one is able to identify and parametrize factors which are mostly conceptual in origin
- that the multiparameter analysis is truly statistically significant, i.e. that one can find out whether or not the variables used are independent
- that the individually parametrized effects are truly additive, i.e. that there is no interaction between the various effects.

Parametrization of Factors of Conceptual Origin. There is much work in the literature on the methods of parametrizing factors of conceptual origin, for both structure effects (14,16), and solvent effects (17,18). An example of this is the decomposition of the parameter σ into σ_I for polar effects, and into σ_R (25) for resonance effects, and the AN and DN (26) scales of the acceptor and donor properties of a solvent. It is not our intent to question the "purity" of such parameters here, but to stress that they are mostly derived by manipulations of analogical parameters. Thus, while methods of analyzing the mechanisms and factors which influence reactivity have gained in qualitative depth, an important part of the raw information in analogical models has at times been lost.

However, from a practical standpoint, these multiparameter models are explicative tools of the reactivity the importance of which is not negligible. For solvent effects, for example, since there are highly diverse solute-solvent interaction mechanisms, it is obvious that single parameter modelling cannot be valid. Palm and Koppel (17) satisfactorily predict a great many effects by using a 4-parameter model: $A = A_o + yY + pP + eE + bB$.

Independence of Microscopic Parameters of Conceptual Origin. Among the work on the independence of microscopic parameters of conceptual origin, there is a fine principal component analysis of the usual solvent parameters (Fig. 8) by Chastrette (27) which provides a good estimate of the effects through the use of only 2 parameters: one for electrophilic reactions (e.g. AN), the other for nucleophilic properties (e.g. DN or E_T). With these two parameters, more than 80% of the solvent effects are correctly described. With an additional parameter, 90% of the data can be estimated.

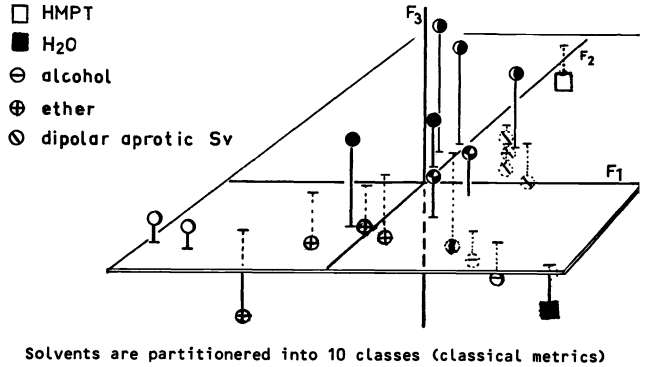


Fig. 8. Solvent Representation in the Space of the Three Main Factors. The names of most characteristic solvents are given. For the other solvents see Ref. 27.

Group Parametrization. One of the most classic methods used for modelling structure effects consists in parametrizing the effects of molecular fragments or groups and considering that the representative parameter of the overall molecular structure is equal to the sum of the parameters of the fragments which constitute it. For example, this is the case for linear free energy relationships based on the polar or steric constants of alkyl groups (16).

$$\log k/k_o = \rho^* \Sigma\sigma^* + \delta \Sigma E_S$$

Therefore, group effects are considered as being additive. In reality, despite the large number of data handled by methods of this type, polysubstituent effects are generally ill-described by group parameters, because of the existence of interactions. In particular, our recent analysis of E_S parameters (28) elegantly demonstrates the importance of these interactions between groups, whether bonded or not. One cannot expect the effect of a group to have a universal value, because this effect might depend on the other fragments of the molecule. In other words, the environment of a group modifies the magnitude of its effects. This fundamental limit of group parametrization is one which is surmounted by the DARC-PELCO method, where group parameters are replaced by site parameters.

Topological Modelling and Site Effects: DARC-PELCO Method. The DARC-PELCO method (29), proposed as of 1967, is one of the correlation methods of the DARC system. It makes it possible to seek a precise relationship between the variation in a structure (S) and its associated information (I), generally with a relatively homogenous population. The DARC-PELCO method is based on the principles of the Topology-Information theory according to which the generation of associated information I is synchronous with the synchronous generation of structure S and hyperstructure HS (Fig. 9).

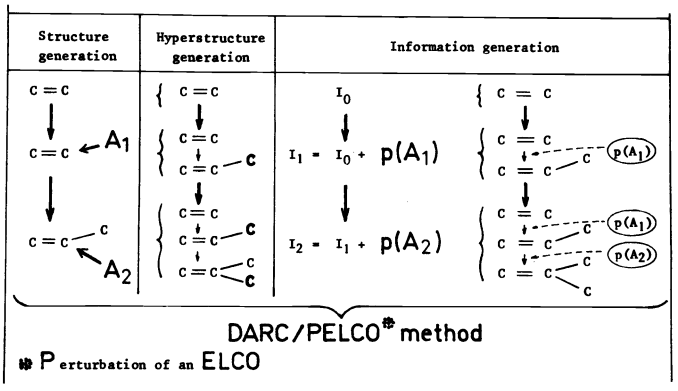


Fig. 9. Synchronous Generation of Structure-Hyperstructure-Information

The following concepts are used to express the synchronism of S/HS/I:

- the population trace, TR(P), a graph that regroups all the ordered sites appearing at least once in the environments of all the structures in P, and characterizes space HS where the structures of P and those predictable from P are located;
- the topochromatic vector $\vec{T}(\&)$, a boolean vector associated with the environment of each structure, that indicates the presence or absence of each ordered site.

This vector provides a quantitative expression of the compound's structure in a population characterized by its trace. Moreover, it directly accounts for the overall topology and chromaticity (nature of bonds and atoms, geometrical and stereochemical data) of the structure described, and retains the local information.

The topochromatic vectors $\vec{T}(\&)$ and all information I associated with each structure are used to establish a Topo-Information relationship whereby the topochromatic variations between structures S in population P are related to the variations in information I for these compounds. Thus:

$$I(\&) = \langle \vec{T}(\&) \mid \vec{I} \rangle_{TR} = \sum_{s \in TR} \bar{p}(s) \cdot \&(s)$$

where I(&) represents the activity associated with environment & in a compound, $\bar{p}(s)$ represents the component of the information vector $\vec{I}$ expressing the average perturbation of activity associated with the introduction of topochromatic site s into &, and &(s) represents the component corresponding to site s in the topochromatic vector $\vec{T}(\&)$.

In the example of the bromination of aliphatic alkenes in methanol, the regression conducted with the focus and topochromatic sites as explicative variables yields highly satisfactory statistical criteria and describes the evolution of the data with the structural modifications with great accuracy (30) (Fig. 10). The field of application and the predictive capacity of this correlation are greater than when using classical methods (Table 3). The proference of the correlation based on 70 compounds is 106 compounds, 37 of which were predicted by retrospective proference (Table 4). This score greatly exceeds the 10 or so proference compounds obtained by classical methods.

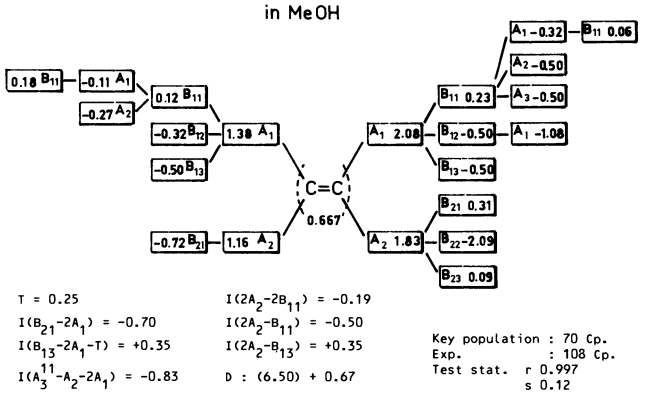
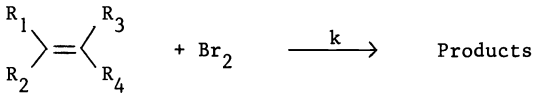


Fig. 10. PELCO Valuated Graph of Alkene Reactivity vis à vis Bromine (log k_{Br_2})

The valuated space groups together the focus and the set of topochromatic sites of a population of 70 olefins over a reactivity range of 6.50 log units. By taking into account the synchronous generation of the structures, the hyperstructure and the information, it is possible to show the contribution from each site to the overall property and the existence of interactions between sites.
The rate constant of a given olefin can be obtained by calculating the sum of the contributions from each individual site (in boxes) and from the interactions between unrelated sites. The valuation here is a preliminary estimate based on a population limited at 70 alkenes.

TABLE 3. Limits of Analogical Correlations in Alkene Bromination



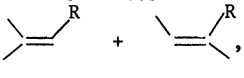
$\rho^*\Sigma\sigma^*$ correlation

$\rho^* = -3.1$ $R = 0.9889$

valid for 37 alkenes with linear substituents

$\rho^*\Sigma\sigma^* + \delta\,\Sigma E_s$ correlation

$\rho^* = -4.3$ $\delta = +0.5$ $R = 0.9873$

valid for 48 alkenes (37 +  + , R = iPr,tBu)

non-correlatable

60 alkenes with t-Bu and neo-Pe substituents + several branched substituents.

Table 4. Performance of Classical and Non-Classical Correlations

Correlation	nb. of compounds in the corr.	nb. of parameters	Coef.	nb. of calculable compounds	Prediction
$\rho\sigma^*$	37	5	0.9889	45	8
$\rho\sigma^* + \delta E_S$	48	16	0.9873	58	10
$\mathscr{E}$	70	21	0.9975	176	106 (37 of which are checked experimentally)

Aside from prediction, the PELCO method, while highlighting certain outstanding facts, provides elements which are precious for the understanding of the phenomenon: the relative value of the perturbations associated with each site, presence of important interactions between certain sites unrelated by the generation. The comparison of the representative valuated graphs established in various solvents, (e.g. CH_3COOH and MeOH) should make it possible to show the influence of the solvent at the level of each site.

Modelling the Effect of all the Variables: Interactive Processes

Using linear free energy relationships when a single structure or solvent effect is involved yields limited, but generally good, empirical models of reactivity. The chemist must choose the empirical parameter scale best suited to the correlation of his data. Which model should be used when several factors intervene simultaneously ?

Independence of Variables: Limitations

In a first approximation, the effects of the various factors are generally considered to be additive, i.e. the interaction term $\phi(M_i, X_j)$ is considered to be nil. For example, if a solvent effect is modelled by a Winstein-Grunwald equation and a structure effect is modelled by a Hammett equation, the sum of both effects yields an overall reactivity model:

$$\log k/k_o = \rho\sigma + mY$$

This hypothesis is based on the assumption that ρ is constant whatever the solvent, and that m is not dependent on the structure. However, experience shows that this is not generally the case and that such a model does not contain all the information. For example, in the bromination of styrenes (31), two highly different values of ρ for the ring substituent effects are measured in methanol (-4.8) and in acetic acid (-6.2) . Therefore, ρ is not constant and implicitly contains information about the solvent effect. Likewise, the value of m measured in the bromination of 1-pentene (23) (1.16) is much higher than that measured for the bromination of isopropenyl acetate (32) (0.63), thereby indicating that m is not constant and that it contains information about the structure effect.

An extensive study of structure and solvent effects recently conducted by Offermann and Vögtle (33) on the selectivity of the radical bromination of toluene yielded data which indicate the simultaneous variation of ρ with the solvent (from -0.47 in HCOOMe to -0.71 in CS_2) and of the solvent effect coefficients proportional to $\frac{1}{n^2}$ (from 2.5 for 4-tert-butyl toluene to 3.93 for 4-nitro toluene) (Fig. 11).

Two conclusions are drawn from these observations: (i) that the large variations of ρ and m observed for the same mechanism show that attributing a mechanism on the basis of a single value of ρ and m is ambiguous from a qualitative standpoint, and (ii) that, from a quantitative standpoint, the interaction term cannot be neglected, and that the function ϕ initially proposed should be integrated into the reactivity model.

Interaction Function

Determining the function ϕ is easy when the structure and solvent effects are modelled by linear relationships. The Taylor series development of reactivity considered as a function of several variables leads to the following equation in which the interaction function ϕ

$$F(X_j, M_i) = F(X_o, M_o) + \alpha_{X_o} M_i + a_{M_o} X_j + Q X_J M_i$$

is proportional to the product of the parameters expressing the structure effect and the solvent effect.

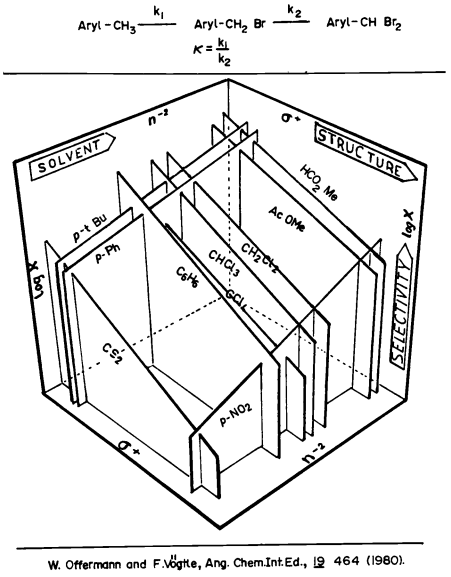


Fig. 11. Non-Independent Structure and Solvent Interaction

When the structure and solvent effects are modelled by Hammett and Winstein relationships, respectively, this equation becomes

$$\log k/k_o = \rho\sigma + mY + q\sigma Y$$

The cross-term coefficient q expresses the sensitivity of ρ to the solvent effect and the sensitivity of m to the structure effect. The initial equation was proposed by Miller (34) and was later developed by Jencks (35) and Dubois (36) for the interaction of several structure effects. Such models of reactivity can actually be applied whatever the type and number of interactions.

Offermann and Vögtle's data (33) can be modelled by the following equation

$$\log X/X_o = 2.81\sigma^+ - 1.31 \frac{1}{n^2} + 1.49\sigma^+ \frac{1}{n^2}$$

This model can be used to accurately predict the selectivity of the bromination of any toluene in any solvent.

Thus by taking these interaction functions into account, the stored information is complete and unambiguous: complete because the reactivity can be calculated with accuracy whatever the case, unambiguous because ρ (or m) can be calculated for all solvents (or structures) and can thereby be associated with a specific mechanism.

The introduction of interaction functions in parametric models makes it possible to improve the prediction as well as the understanding of the reactivity. This improvement is acquired by a significant experimental effort since setting up the networks corresponding to the multiparameter functions requires a much larger number of multifactorial data than the acquisition of classical monolinear functions. However, only the treatment of vast sets of data related to large reactivity ranges makes it possible to grasp a significant and unambiguous part of the physicochemical reality.

The introduction of cross-terms in linear free energy relationships is a means of working in a larger resolution space which makes it possible to discern terms whose physical significance must be specified.

A feature of the DARC-PELCO method for treating data is that it deals with families of data which are much vaster than those usually handled, and thereby provides means for solving multiple reference situations where complex interaction terms are also involved. Thus, from a set of DARC-PELCO valuated graphs for the bromination of alkenes, it is possible to obtain estimates of solvent effects and structure (geometry, conformation) effects either by studying the correlation data or by using standard external parameters.

An overall method of modelling, which supposes non-negligible means of calculation, will little by little make it possible to conciliate the understanding of a general type of

reaction with more rigourous though more restrictive elucidations in certain well-defined areas.

OUTLOOK FOR THE FUTURE

Whereas the initial aim of reactivity modelling was to understand physicochemical phenomena while predicting them, the trend over the last few years has been to put prediction before understanding. Indeed, the change from single parameter analyses to multiparameter correlations, which has been made easier by the application of statistical techniques to chemistry, makes it possible to account for large sets of data without necessarily conferring more specific physicochemical significance on these treatments than the first analogical models. By refining multiparameter models, whether of a mathematical or physicochemical nature, it is possible to obtain representations of available data which are precise and which permit fairly reliable predictions, at least by interpolation. However, at the interpretation level, there is a growing awareness that the impact of these methods on the understanding of the reaction process is relatively restricted. Thus, although multiparameter methods have broadened the useful range of correlation analysis, they themselves have their limits, particularly where the understanding of reaction mechanisms and the appreciation of the factors which determine reactivity are concerned.

Actually, the partial failure of current analyses stems essentially from the need for modelling. Indeed, whatever the principle, every modelling method locks the dynamic reality in systems rendered rigid by constraints hinging on the necessary schematization of the effects involved. Existing models are usually too restrictive and necessarily limit the number of degrees of freedom of true reaction systems. Therefore, it is now a matter of going beyond the present notion of modelling in order to be free of its defects. However, since methodologies for handling the phenomena generally force one to rely on a logical structuring of data, the chemist's current aims to reach a true understanding of chemical dynamics consist in elaborating new models which are essentially flexible and based on clearly defined hypotheses.

In conclusion, there is a real need to define methodologies oriented towards chemical problem solving where computing services are used to advantage.
Their advantages lie in:
a) their being based on efficient handling potential of all chemical data (physical and structural),
b) their relying on a set of solution strategies programmed so as to represent the best expert knowledge and
c) their making possible an interactive relation with the chemist during the iterative resolution of specific problems.

The time has come to extend our ideas of modelling, which have too long imprisoned the chemist within rigid boundaries unknown to chemistry itself. The tools must become worthy of the art.

Acknowledgment - The author wishes to express his appreciation to his co-workers, whose names appear in the references.

REFERENCES

1. DARC : Description, Acquisition, Retrieval, Computer-aided design.
2. a) J.E. Dubois, "The Chemical Applications of Graph Theory" Ed. A.T. Balaban, Academic Press. 1976, pp. 330-376;
   b) J.E. Dubois, "Computer Representation and Manipulation of Chemical Information" Copyright 1974 by W.T. Wipke, S. Heller, R. Feldmann and E. Hyde;
   c) J.E. Dubois, Israel Journal of Chemistry, 14, 17 (1975).
3. T. Kitagawa, VIIIth International Congress on Cybernetics, Namur, 1976.
4. For a review of the principle systems, see: Computer-Assisted Organic Synthesis, Eds., W.T. Wipke and W.J. Howe, American Chemical Society, Washington, D.C., 1977; Computer-Assisted Structure Elucidation, Ed., D.H. Smith, American Chemical Society, Washington, D.C., 1977.
5. a) J.E. Dubois, D. Laurent et H. Viellard, C.R.Acad.Sci., 263C, 764 (1966);
   b) J.E. Dubois, D. Laurent et H. Viellard, ibid., 263C, 1245 (1966);
   c) J.E. Dubois, D. Laurent et H. Viellard, ibid., 264C, 348 (1967).
6. a) J.E. Dubois et H. Viellard, Bull.Soc.Chim., 900 (1968); ibid., 905 (1968); ibid., 913 (1968); ibid., 839 (1971);
   b) J.E. Dubois, A. Panaye et P. Cayzergues, C.R.Acad.Sci., 290C, 429 (1980);
   c) P. Cayzergues, A. Panaye et J.E. Dubois, C.R.Acad.Sci., 290C, 441 (1980).
7. J.E. Dubois, D. Laurent, A. Panaye et Y. Sobel, C.R.Acad.Sci., 280C, 851 (1975); ibid., 281C, 687 (1975).
8. IGLOO: Invariant Graph and Localized Ordered Operators, J.E. Dubois et al. to be published.

9. The definition of a common sub-structure is a delicate problem. See: M.M. Core, R. Venkataraghavan and F.W. McLafferty, J.Amer.Chem.Soc., 99, 7668 (1977); T.H. Varkony, Y. Shiloach and D.H. Smith, J.Chem.Inf.Comput.Sci., 19, 104 (1979); J.E. Dubois et al. to be published.
10. PARIS: Pertinent Analysis of Reactions by IGLOO Simulation, J.E. Dubois et al. to be published.
11. J.E. Dubois, A. Panaye and C. Lion, to be published.
12. M. Sjöström and S. Wolde, "Correlation Analysis in Chemistry", N.B. Chapman and J. Shorter Ed., Plenum Press, London, p. 439 (1978).
13. O. Exner, ibid, p. 439.
14. J. Shorter, ibid, p. 119.
15. O. Exner, "Advances in Linear Free Energy Relationships" N.B. Chapman and J. Shorter Ed., Plenum Press, London, p. 1 (1972).
16. J. Shorter, ibid, p. 71.
17. I.A. Koppel and V.A. Palm, ibid, p. 203.
18. C. Reichardt, "Solvent Effects in Organic Chemistry", Verlag Chemie, Weinheim (1979).
19. M. Charton, Prog.Phys.Org.Chem., 10, 81 (1973).
20. J.G. Kirkwood, J.Chem.Phys., 2, 351 (1934).
21. A.J. Parker, Chem.Rev., 69, 1 (1969).
22. J.E. Dubois, M.F. Ruasse and A. Argile, Tetrahedron, 31, 2921 (1975).
23. F. Garnier and J.E. Dubois, Bull.Soc.Chim.Fr., 3797 (1968).
24. M.F. Ruasse and J.E. Dubois, J.Amer.Chem.Soc., 97, 1977 (1975).
25. R.W. Taft and J.C. Lewis, J.Amer.Chem.Soc., 81, 5343 (1959).
26. V. Gutmann, Electrochimica Acta, 21, 661 (1976).
27. M. Chastrette, Tetrahedron, 35, 1441 (1979).
28. J.E. Dubois, J.A. MacPhee and A. Panaye, Tetrahedron, 34, 3553 (1978); ibid, 36, 759 (1980); ibid, 36, 919 (1980).
29. J.E. Dubois, D. Laurent et H. Viellard, C.R.Acad.Sci., 264C, 1019 (1976); J.E. Dubois, D. Laurent et A. Aranda, J.Chim.Phys., n° 11-12, 1608, 1616 (1973).
30. J.E. Dubois, unpublished results.
31. M.F. Ruasse, A. Argile and J.E. Dubois, J.Amer.Chem.Soc., 100, 7645 (1979).
32. E. Bienvenue-Goëtz and J.E. Dubois, J.Chem.Res., (M) 2249 (1979).
33. W. Offerman and F. Vögtle, Angew.Chem.Int.Ed., 19, 464 (1980).
34. S.I. Miller, J.Amer.Chem.Soc., 81, 101 (1959).
35. E.M. Cordes and W.P. Jencks, ibid, 84, 4319 (1962).
36. a) J.E. Dubois, J.J. Aaron, P. Alcais, J.P. Doucet, F. Rothenberg and R. Uzan, J.Amer.Chem.Soc., 94, 6823 (1972);
b) Recent extensions on these multi-substituents effects (MSE) correlations with cross-interactions terms in generalized Hammed relations, with M.F. Ruasse and A. Argile, will be published soon;
c) A. Argile, Doctoral Thesis, University of Paris VII, 1980.