Tetrahedron Vol. 49, No. 16, pp. 3367-3386, 1993
Printed in Great Britain

0040-4020/93 $6.00+.00
Pergamon Press Ltd

Similar Group Interferences. A General Approach to the Location of Interfering Functionalities.

Luca Baumer, Giordano Sala, Guido Sello*.

Dipartimento di Chimica Organica e Industriale, Universita' di Milano, via Venezian 21, Milano, Italy.

(*Received in UK 7 May* 1992)

Abstract. The location and evaluation of interfering functionalities are aspects of great importance in the synthesis planning, because the success of a synthesis often depends on the correct management of functionalities involved both in the formation of the strategic bond and in the multiplication of the reactive sites. The present approach introduces a new subdivision of interferences acting either locally (Similar Group Interferences) or generally (Group Environment Interferences), and suggests a solution for the first case. Using a group of molecular descriptors the level of local interference of a functional group is calculated and the possibility to be an alternative reacting site is determined. Electrophilic and nucleophilic situations are separately considered and the complete procedure for the identification of SGIs is described. Example studies are discussed in both simple and complex structures.

INTRODUCTION

The location and evaluation of interferences caused by functional groups of an organic molecule represent a puzzle whose solution has been a long standing problem for the synthetic chemist. In fact the chance of eliminating useless transformations, as for example protection / deprotection of functional groups, can give such an important practical saving in real terms that a particular synthetic route can be preferred to other much more elegant ones.

If the need for rationality, required by theoretical models applied to organic synthesis planning, is added to the purely synthetic aspect, a scenario is obtained in which consideration of interferences is a really fundamental part. In spite of this, in specific literature there are few solutions that can, at the same time, both solve the need to get a complete view of the problem and furnish good quality results.

The problem of the location of interferences can be explicitly or implicitly approached. In the first case, having located the key transformation of the synthetic step, the molecular substructures, which can be modified somewhat by the considered transformation are sought. The discovered substructures are signalled and, if possible, a particular protection is suggested. In the second case, the interference appears only when the transformation is applied, giving unwanted results. The best-known example in the first class is represented by the LHASA program[1,2] and, more generally, by all empirical programs; whereas the second class includes all the mechanistic programs (SYNGEN, EROS, etc.)[3-5].

In an attempt to approach and solve this problem, using a standardized methodology that could be used inside a program for the synthesis planning (Lilith)[6] presently being studied in our department, our first step has been to correctly define the limits and characteristics of the problem. The best division of interferences has been sought distinguishing two aspects: the Similar Group Interferences (SGI)(for a complete list of abbreviations see Appendix 1), that regard the interactions in the behavior of comparable functions and whose solution requires the accurate choice of reaction partners and/or protective steps, and the Group Environment Interferences (GEI), that include all the transformations brought about by the

reaction conditions.

As the GEIs are caused by a group of factors that are seldom microscopically well-known and as a tentative general classification of them does not even exist, it was preferred to examine at first the SGIs that, besides being a molecular aspect more easily classifiable, are the step which immediately follows reactivity evaluation in the logical flow of our program.

A schematic description of the working style of the Lilith program is required for a better understanding of our choice. The program is independent from a data base; but it uses a group of physical and topological descriptors of the target molecules to perform some tactical and strategical choices, quantifying their values. The program sections achieved up until now are: 1) the choice of the sets of strategical bonds[7,8]; 2) the definition of the transformations that can permit their synthesis[9]; 3) the evaluation of the SGIs.

The search and location of SGIs are strictly related to the nature of the key transformation, which, in Lilith, is determined by the electronic situation present on the target molecule. As the program explicitly considers only heterolytic reactions, that is electrophile / nucleophile interactions, the SGIs regard solely the presence of molecular positions with electrophilic / nucleophilic characteristics similar to those of the atoms directly involved in the transformation.[10]

Positive and negative reacting atoms

In the majority of known cases, it is possible to observe a different behavior of the atoms of various signs. For the negative atoms it is possible both to make preformed ions (in the case of intermolecular reactions), and to choose the reaction conditions so as to render non-interfering those positions that are less reactive or much more reactive. Vice versa this possibility is seldom present for the positive atoms. Therefore a scheme is obtained for the positive atoms where the SGIs are considered with all the positions of similar or stronger reactivity, both intramolecular and intermolecular; while only intramolecular SGIs with positions of similar reactivity exist for the negative atoms. The objections against this simplification can be of two kinds. Regarding the intramolecularity, it is possible to foresee situations where the preformed anion cannot avoid the interchange with a position so reactive to be insensitive to the classical neutralization methods (rate of the reagent addition, inversion of the addition order), but this situation is quite unusual and can often be overcome. On the other hand, the presence of many more reactive positions can be source of difficulties in the choice of the reaction conditions and can thus invalidate the analysis; however, these positions are identified and can be, at will, inserted into the SGIs.

The possibility of distinguishing two different reactional situations for the positive atoms is offered by different reaction mechanisms: the presence of a tetrahedral reacting position, or of a trigonal or digonal reacting position. The molecular characteristics that control the two situations are, in fact, quite different. This difference, on the contrary, is not present for the negative atoms.

In substitution reactions the reaction rate is determined, forgetting for the moment the geometrical factors, by the nucleophilicity of the entering group and by the nucleofugacity of the leaving group. In an attempt to render the behavior of the two reaction partners independent, only the second factor deserves examining.

On the contrary the electrophilicity of the positive atom is interesting in the case of addition reactions. It is clear that the molecular features that determine the nucleofugacity level of a group are different from those controlling its electrophilicity.

Negative atom reactivity is, on the other hand, connected to the atoms' nucleophilicity and/or basicity. In this case it is however useless to divide up the analysis.

PROCEDURE DESCRIPTION

The activity of location and evaluation of the SGIs is done using a procedure that is in turn divided up into different subprocedures devoted to diverse activities. The whole process can be set out in four sections:

1) a main section (called INTERF) that takes care of the distinction of some fundamental situations and of the distribution of the work to the subordinate sections; 2) a section that updates the representation of the molecular structure (called MKPREC); 3) a section that analyses the negative atom cases (called ACCDEN); 4) a section that deals with positive atoms (called SMANIA).

INTERF

The solutions proposed by Lilith are composed of a group of bond breaks of the target molecule that permit it to be divided up into precursors[7]. Each solution is the input to the procedure, together with the information connected to the polarity preferred by each bond[9]. The first INTERF operation is the transformation of the target into two precursors that have, in the place of the broken bonds, groups that can simulate the desired behavior. This operation is performed by MKPREC.

Having obtained the precursor molecules, it is necessary to separate the different reactivities, that is to distinguish among the three possible alternatives: a) negative atom (ANION); b) positive atom involved in a substitution (LEAVING); c) positive atom involved in an addition (ATTACK).

ANION. First of all the routine that calculates the nucleophilicity or the acidity of the corresponding proton (ACCDEN) is called for each negative atom. The use of only one routine is possible because the reactivity depends on the same factors. It is however clear that the reactivity is not comparable for atoms belonging to the two groups, as is after all inherent to the SGI definition.

In fact a separation of the two cases is carried out as soon as the reactivity factor has been calculated, distinguishing the atoms that must undergo deprotonation from the others. In the first case the ACCDEN routine is used to evaluate the reactivity of all the protons of the molecule; in the second case the potential nucleophilicity is calculated only for atoms of the same type as ANION. The calculation is also limited to atoms that are in the molecule containing ANION. Only those atoms that have a reactivity similar to that of ANION are signalled as interfering.

ATTACK. Also in the case of ATTACK, a distinction is first made between intermolecular interactions (where the interferences on the fragment containing ATTACK (called real interferences) are divided from those on the fragment containing ANION (called selfcondensations)) and intramolecular (where selfcondensations and real interferences are indistinguishable). The identification procedure calls for a first provisional assignment of interference to all the atoms that are involved in an activated multiple bond (i.e. where one of the spanning atoms is sufficiently less electronegative than the other); the atoms in aromatic rings are then eliminated (e.g. the nitrogen in a pyridine). Moreover, if ATTACK is involved in an addition-substitution reaction, the interferences are maintained only for similar situations; the presence of such a case is determined by using ALPHAHardness (ALPHAH), that must have similar values for ATTACK and for the interfering atom.[11]

Finally, if ATTACK is not activated towards addition reactions (and for some logical[9] reasons has nevertheless been defined positive) the interferences are not considered.

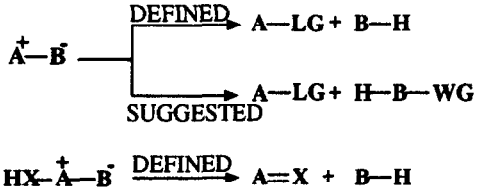
LEAVING. In this case the nucleofugacity of LEAVING is first evaluated followed by that of the other atoms which have a chance to interfere. Two particular aspects must be emphasized: 1) only the nucleofugacity of the best leaving group present on an atom is evaluated, both for LEAVING and for the others; 2) only bimolecular nucleophilic substitution reactions (SN_2) are considered. In this perspective, after calculating the nucleofugacity by using the SMANIA routine, the following are eliminated from the SGI group: 1) those atoms too hindered to sustain an SN_2; 2) those atoms that are bridgeheads or belong to aromatic rings; 3) the halogen atoms that have been inserted by MKPREC to be transformed into organometals and that, even if theoretically interfering, cannot be classified as SGIs.

MKPREC

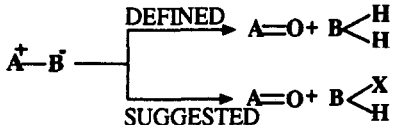
The structures of the precursor molecules, that is passed to the section examining the SGIs, come from the directly preceding part that determines the reactivity present on the target; at this level the structures differ from the target only because of the breaking of the determined strategic bonds. Therefore it is necessary to better specify them so as to be able to evaluate more precisely the reactivity levels and to compare them with the remaining part of the molecule.

MKPREC takes care of the introduction into the precursors of the opportune substituents on the atoms spanning the bonds which have just broken. Moreover it makes the further changes to increase those reactivities not sufficiently specified.

Three main situations are distinguished (see Schemes 1-3): 1) the break of only one bond of the first order; 2) the break of only one bond of the second order; 3) the break of two bonds with an atom in common.



Scheme 1. The structural changes of the precursors in the case of the break of a single first order bond. LG is a generic leaving group, WG is a generic electron withdrawing group, X is a heteroatom. DEFINED and SUGGESTED are levels of atom polarity.

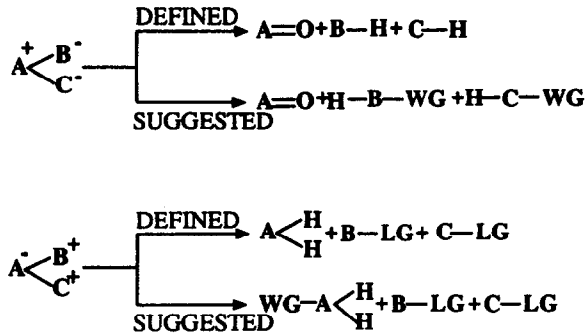


Scheme 2. The structural changes of the precursors in the case of the break of a second order bond. X is a generic heteroatom. DEFINED and SUGGESTED are levels of atom polarity.

Lilith determines two levels of activation (DEFINED and SUGGESTED) for the atom polarities[9] (both negative and positive), plus an undetermined case (UNDEFINED), where the atom polarities cannot be assigned. In this case the analysis is similar to that carried on for the SUGGESTED case. In fact the UNDEFINED solutions are transformed into two results with opposite polarities, with inversion on each atom and with activation at the SUGGESTED level. The structural updating is therefore included in the following.

Different structural modifications are made depending on the polarity of the involved atom and on its activation level.

1) Positive polarity. For any activation level a Leaving Group (LG) is attached to the atom, unless a heteroatom carrying a hydrogen atom is positioned alpha to the reacting atom. In this case the hydrogen atom is eliminated and the bond order between the heteroatom and the reacting atom is increased (e.g. an alcoholic group is transformed into a carbonyl). As the reaction that will really come about is still too undefined, it is impossible to determine the exact nature of the LG to use and thus a generic hypothetical LG is introduced that changes as little as possible the electronic situation of the reaction center and that is, at the same time, a good LG; in the present case an iodine atom is used.



Scheme 3. The structural changes of the precursors in the case of the break of two geminal first order bonds. LG is a generic leaving group, WG is a generic electron withdrawing group. DEFINED and SUGGESTED are levels of atom polarity.

Negative polarity. In the case of a DEFINED activation level, a hydrogen atom is introduced, because a deprotonation is directly applicable. For the SUGGESTED level two cases are distinguished: a) the definition derives from molecular electronics. If the reacting atom is unsaturated a halogen is introduced foreseeing a metallation; if it is saturated a further activation is tentatively introduced either by changing an already existing group into an activating group, or by introducing a new activating group (represented by a carbonyl). b) The definition comes from an umpolung operation; in this case a hydrogen atom is added because the umpolung reaction must greatly change the structure and, at this stage, it is impossible to foresee a correct analysis of the SGIs.

2) Positive polarity. At any definition level an oxygen atom is introduced which is double bonded to the reacting atom representative of the groups with activated multiple bonds.

Negative polarity. In the case of an activation of DEFINED level two hydrogen atoms are added, envisaging an addition-elimination reaction. In the case of an activation at the SUGGESTED level one hydrogen atom and one halogen atom are introduced, to represent the ylide preparation reaction.

3) Positive polarity. At any definition level an oxygen atom, doubly bonded to the reacting atom, is introduced; in this case the hypothesis regards addition-substitution and not addition-elimination reactions. The presence of the oxygen permits the activation for the addition and also permits easy transformation into an LG for the substitution.

Negative polarity. In the case of DEFINED level activation, two hydrogen atoms are introduced, inferring two subsequent deprotonations, both already activated. In the case of SUGGESTED level activation, research similar to that carried out in case (1) is performed, trying to activate the examined position.

ACCDEN

The ACCDEN procedure is involved in the evaluation of the problems that can be present at the time of preparing the atom with nucleophilic activity. These problems are essentially of two kinds depending on the type of the atom involved; if the atom needs to be transformed into an anion to become nucleophilic, it is necessary to examine the deprotonation step; if, to the contrary, the atom is nucleophilic in the neutral state, the calculation of its activation level is required. An aspect of the differentiation is inherent to the choice of the atoms used in the comparison (hydrogen atoms or heteroatoms of the same type), whereas a second aspect is calculated by the ACCDEN procedure.

There are two main factors used by the procedure to evaluate the reactivities: 1) the electronegativity level; 2) the accessible conjugations. These two main factors are then corrected using the number of the activating groups and the presence of inductive activation, the latter only if the conjugation is absent. A final

classification into groups, to improve the analysis standardization, and the location of the atoms belonging to the same group, and therefore interfering, conclude the analysis.

1) The first factor is the activation level compared with the reference situation of a non activated tetrahedral carbon. This factor makes available the first data which signal both the tendency of the atom to deprotonate and the level of its potential nucleophilicity.

$$F_1 = (\chi_{ANION} - PSUB) \times PQ$$

where χ is the calculated electronegativity of ANION, PSUB is equal to 2.51, the electronegativity of sp^3 carbon, and PQ is a multiplicative factor equal to 3000 to scale the difference.

2) The capacity of delocalization on a conjugated system of both the charge coming from the deprotonation and the electronic richness useful for a nucleophilic attack determines an increment or a decrement of the reactivity that must be considered. By using again the electronegativity differences and the distance, in terms of multiple conjugated bonds, from the reaction center, it is possible to estimate the relative stabilization of the atom.

The factor is calculated for all the activating groups. If many activating groups are present their contributions are not simply added together (it is in fact well known that the described property is not additive), but they are scaled down using the best contribution first.

$$F_2 = \Sigma_i CONJPL_i \times P1$$

where P1 is a scaling factor equal to 1000, the sum is done on all the activating groups α to the reacting atom that can conjugate, and

$$CONJPL_i = \Sigma_k (CONJP_k - CONJP_i) / (CONJP_k + CONJP_i)$$

where the sum is done on all the groups (k) more activating than group (i)

$$CONJP_i = \Sigma_j (PCON + \chi_j - PSUB) / (PCON \times DECAY)$$

where PCON is equal to 4, PSUB is equal to 2.51, the sum is done on all the atoms interested by the conjugative mechanism in each group, and

$$DECAY = BO \times BO \times (1 + CUTEN)$$

where BOxBO is a negative factor equal to the distance of the considered atom from the reacting atom in terms of conjugations, and CUTEN is equal to 0.35 and takes care of a standard contribution given by a single direct conjugation.

If, and only if, multiple bonded systems are absent, the inductive effects caused by the presence of groups with different electronegativity are considered. They are also non additive and therefore a system using scaled contributions is used.

$$F_2 = \Sigma_i PDEN \times P(G) \times (\chi_i - PSUB)$$

where PDEN is a constant scaling factor equal to 0.6, P(G) is a variable scaling factor equal to 1000, 750, or 400, depending on G that is the number of the activating groups present on the reacting atom, PSUB is equal to 2.51.

Then F_1 and F_2 are summed giving the activation factor

$$F = F_1 + F_2$$

The final operation is the classification of the reaction center and of its possible SGIs in activation groups. This device is not simply an ordering factor, but it permits the overcoming of the too rigid classification connected with calculation of exact numerical quantities. It is thus possible to ensure a better consideration of the generality of the possible structural situations. Six groups are used that correspond, qualitatively, to an activation: 1) null ($F < 300$); 2) weak ($300 < F < 700$); 3) like an alcoholic proton ($700 < F < 1000$); 4) like an alpha carbonylic proton ($1000 < F < 2000$); 5) like a phenolic proton, or a proton on carbon that has at least two activating groups ($2000 < F < 3600$); 6) like a carboxylic proton ($F > 3600$). It is clear that the proposed ordering is reversed for the evaluation of the potential nucleophilicity, where all the mechanisms that favour the dispersion of the electron density disfavour the nucleophilicity; after all, the evaluation of the SGIs only requires the membership to the same group and the absolute level of activation is unimportant.

The SGI determination is done by signalling all the atoms inside the activation group of the reacting

atom with the addition, in the case of proximity to one of the group borders, of the atoms with similar activation. The comparison ranges are naturally different, larger (between 400 and 1600) for membership of the same group and very small (150) for the intergroup SGIs.

SMANIA

The Leaving Group class is the last to be analyzed. The evaluation of the nucleofugacity of a group has been discussed by many chemists[12-17] and is still lacking a well defined approach that can distinguish, at a very reliable level, the unlimited possible alternatives. The proposed procedure is not aimed at providing the so long sought-after solution; on the contrary, it must be interpreted as the best way that has been found for solving the narrower SGI problem.

The main factor that characterizes the degrees of nucleofugacity is the local hardness[18] difference between the reaction center and the LG, i.e. the contribution to unbalancing the hardness between the reacting atom and the LG caused by the molecular neighbourhood (called AlphaHardness Delta or DELTAH).

In a first approach the molecular neighborhood was defined by the atom alpha sphere[19]; the present version, on the other hand, extends it to all the atoms electronically interacting with the considered atom, that is belonging to the same functional group (with a recently introduced meaning[20]).

$$DELTAH = |ALPHAH(CENTRE) - ALPHAH(LG)|$$

where ALPHAH(CENTRE) is the local hardness of the reacting atom and ALPHAH(LG) is the local hardness of the leaving group.

In the majority of the analyzed cases, the DELTAH can alone characterize the nucleofugacity. There are corrections in relation to other molecular features; they are essentially three: 1) the electronegativity difference between the reaction center and the groups different from the LG, corrected by the bond order; 2) the LG electronegativity; 3) the presence of ring strains.

1) In this case, the changes given by the presence of groups (NG) that strengthen or weaken the bond between the reaction center and the LG are considered, keeping in mind that, depending on the electron withdrawal or donation, the presence of multiple bonds respectively increases or decreases the LG nucleofugacity.

$$DELTAH = DELTAH \times BORD(CENTRE, NG)$$

if NG is more electronegative than CENTRE, or

$$DELTAH = DELTAH / BORD(CENTRE, NG)$$

if CENTRE is more electronegative than NG. BORD is the order of the bond between CENTRE and NG.

2) Two situations are distinguishable also for the LG electronegativity: when the LG is very electronegative (compared with an unsubstituted atom of the same type) its nucleofugacity is not greatly influenced, whilst in the case of scarce activation the DELTAH must be corrected, because the electronegativity increases the LG nucleofugacity. In other words, since the ALPHAH can be considered as the steepness of the chemical potential change, in a situation where this steepness is high, the electronegativity (i.e. the corresponding starting height) is unimportant, whilst it is relevant with less steep slopes. Therefore in the second case

$$SMANIA = -(DELTAH - \chi(LG)) \times DELTAH$$

where SMANIA is the nucleofugacity and χ is the electronegativity of LG.

3) The ring strain of small rings makes them always more reactive and represents a well-known instability factor. This correction allows, for example, the distinction between an alkoxide and an epoxide LG. A true geometric treatment (still missing in Lilith) would avoid the use of the ring strain as a special factor in the nucleofugacity evaluation that would thus only remain exclusively related to the molecular electronics.

The correction introduced is

$$SMANIA = SMANIA \times 8 / (SIZE(RING) - 2)$$

where SIZE(RING) is the ring size and the correction is applied only if it is less than 5.

RESULTS AND DISCUSSION

The introduction and the discussion of the results will be divided into two main sections. In the first section the data related to the evaluation of each reactivity component (acidity / nucleophilicity, nucleofugacity / electrophilicity) will be analyzed. In the second section the SGIs found by the procedure in molecules of medium complexity subjected to the synthetic analysis of Lilith will be presented. It is obvious and must be greatly emphasized that this second result aspect is the only really important one for the solution of the problem which lies behind the present work. New acidity or nucleofugacity scales are not therefore sought.

Table 1 gives some examples of acidity level calculation carried out by the ACCDEN routine together with the estimated corresponding pK_a taken from different literature sources[21-34]. The molecules used have been chosen following two requirements: that of the existence of an evaluation of their corresponding pK_a (possibly from different sources) and that of their connection with the SGI calculation. Therefore molecules containing unusual or inorganic groups have been excluded and molecules with acidic carbons have been privileged. In the table is also reported the acid group to which the molecule belongs, as this is the main weighing up factor used by INTERF. No attempt has been made to quantitatively correlate the obtained results with those coming from the literature, the discussion will thus only be qualitative.

The literature data come from experiences in three different reference situations: acidity measured in water, in DMSO and in the gas phase. The differences among the reported values have already been thoroughly discussed[26,30] and they will not be considered in any greater depth, except to note the lack of any kind of qualitative or quantitative agreement amongst them. An interesting recent work of Taft and Bordwell[30] remarks on the acidity of some compounds measured in the gas phase and in DMSO; it is interesting to note the structural factors that the authors think influence the acidity. These are: a) charge on HA; b) nature of A; c) charge delocalization on A^-; d) presence of lone pair-lone pair repulsion near A^-; e) presence of phenyl groups; f) ring closure; g) presence of benzofusions; h) presence of dipoles more or less near to the acidic group; i) steric hindrance. The number and variety of the reported factors are impressive; nevertheless each of them seems to have its exact role. If other external sources of perturbation are also considered (e.g. solvation, specific coordination, etc.), it is easily understandable that any general theoretical procedure for the acidity calculation must be very complicated. Therefore the calculation procedure used only takes care of some factors within a semiquantitative scheme to solve our problem. Similar considerations are valid regarding the estimation of the heteroatom nucleophilicity; a problem which is however very simplified by the imposed limitation of interferences only among atoms of the same type and which is less important in synthesis planning which studies carbon-carbon bond formation.

When looking at the calculated values the following can be noted: a) the simultaneous presence of many activating groups is correctly reproduced; b) molecules with acidic atoms and acidifying effects of the same kind are well classified from a qualitative, and often also quantitative, point of view; c) some effects, such as the presence of dipoles and/or polarizable atoms, are underestimated; others, such as the solvent and/or the geometric effects, are completely ignored.

Let us consider some examples in greater detail.

Table 1. Calculated Acidity in Comparison with Measured Data.

Molecule	ACCDEN	Acidity Group	pk_a[b]	Gas Phase[a]
C_4H_9SH	290	Null	11[c]	1/18
CH_3NO_2	3037	Phenol	17.2(10)[c]	2/18
$CHCl_3$	526	Weak	-	3/18
CH_3COCH_3	1360	Carbonyl	35.1(20)[c]	4/18
CH_3CN	1310	Carbonyl	31.3(25)[c]	5/18
CH_2Cl_2	386	Weak	-	6/18
CH_3SOCH_3	1367	Carbonyl	31.1	7/18
HCCH	1324	Carbonyl	28.8(25)[c]	8/18
tBuOH	2211	Phenol	17[c]	9/18
iPrOH	2204	Phenol	16.5[c]	10/18
C_2H_5OH	2188	Phenol	16[c]	11/18
CH_3OH	2183	Phenol	15.2[c]	12/18
$PhCH_3$	1259	Carbonyl	44 (41)[c]	13/18
H_2O	2339	Phenol	15.7[c]	14/18
PhH	340	Weak	43[c]	15/18
NH_3	1062	Carbonyl	40 (38)[c]	16/18
C_2H_4	361	Weak	44[c]	17/18
CH_4	-78	Null	70 (48)[c]	18/18
$CH_3SO_2NH_2$	3309	Phenol	17.5	
CH_3COOH	3041	Phenol	4.5[d]	
$CH_3SO_2CH_3$	2741	Phenol	23[e]	
PhOH	2410	Phenol	9.2[d]	
$HCONH_2$	2180	Phenol	23.5	
$PhCONH_2$	2103	Phenol	23.5[f]	
pNO_2PhNH_2	1840	Carbonyl	21	
HCN	1819	Carbonyl	-	
$PhNH_2$	1798	Carbonyl	30.7	
CH_3COOR	1381	Carbonyl	24[d]	
PhC_2H	1315	Carbonyl	28.8	
pNO_2PhCH_3	1304	Carbonyl	20.5	
$PhCOCH_3$	1275	Carbonyl	24.7	
CH_3COO^-	1223	Carbonyl	-	
$PhSCH_3$	-9	Null	49	
$(NO_2)_2CH_2$	5912	Acid	4.1[e]	
$PhSO_2CH_2NO_2$	5538	Acid	7.1	
$(PhSO_2)_2CH_2$	5140	Acid	12.1	

[a]Gas phase data are given in qualitative order. Ref. 34 [b]Ref. 26. [c]Ref. 34. [d]Ref. 33. [e]Ref. 21. [f]Ref. 32.

Table 1(contd.). Calculated Acidity in Comparison with Measured Data.

Molecule	ACCDEN	Acidity Group	pk_a[b]	Gas Phase[a]
NO_2CH_2COOR	4539	Acid	9.2	
$NO_2CH_2CHCH_2$	3905	Acid	11.1	
$(COOCH_3)_2CH_2$	2784	Phenol	15.7	
$(CH_3CO)_2CH_2$	2749	Phenol	13.4 (9)[d]	
$(CN)_2CH_2$	2616	Phenol	11[d]	
$PhCH_2CN$	2570	Phenol	21.9	
$(Ph)_2CH_2$	2555	Phenol	32.2	
F_3CCH_2OH	2193	Phenol	22.8	
$PhSCH_2CN$	1355	Carbonyl	20.8	
CH_2F_2	1211	Carbonyl	-	
CH_2Br_2	188	Null	-	
$(CH_3CO)_3CH$	4075	Acid	-	
$(Ph)_3CH$	3806	Acid	30.6	
CHF_3	1583	Carbonyl	-	
$CHBr_3$	263	Null	-	

[a]Gas phase data are given in qualitative order. Ref. 34 [b]Ref. 26. [c]Ref. 34. [d]Ref. 33. [e]Ref. 21. [f]Ref. 32.

1) CH_3NO_2, CH_3COCH_3, CH_3CN, CH_3SOCH_3, HCCH, $PhCH_3$, PhH, C_2H_4, CH_4.

The calculated acidity values are, respectively,: 3037, 1360, 1310, 1367, 1324, 1259, 340, 361, -78; the gas phase acidity order is: 2, 4, 5, 7, 8, 13, 15, 17, 18; the pK_a in DMSO are: 17, 35, 31, 31, 29, 44, 43, 44, 70. It can be noted that, excluding $PhCH_3$ (underlined values), the calculation gives results in good agreement with both the experimental sources, with special care to the grouping of the compounds (all the carbonyl-like compounds are grouped together, as all the hydrocarbon-like compounds. Nitromethane is well distinguished)

2) CH_3NO_2, $CH_2(NO_2)_2$; CH_3COCH_3, $CH_2(COCH_3)_2$, $CH(COCH_3)_3$; CH_3CN, $CH_2(CN)_2$; CH_3Ph, $CH_2(Ph)_2$, $CH(Ph)_3$.

The corresponding calculated acidity values are: 3037, 5912; 1360, 2749, 4075; 1310, 2610; 1259, 2555, 3806; the pK_a in DMSO are: 17, 4.1; 31, 13, n.a.; 35, 11; 44, 32, 30. It is possible to note that the introduction of a second activating group is always more important than the introduction of the third one and that the double activation often furnishes a double acidity power, in qualitative agreement with the experimental data.

3) H_2O, CH_3OH, iPrOH, tBuOH.

The corresponding calculated acidity values are: 2339, 2183, 2204, 2211; the gas phase order is: 14, 12, 10, 9; the pK_a in water are: 15.7, 15, 16.5, 17. The calculated order is comparable with the gas phase order except for H_2O that is more acidic; H_2O is overestimated because the electronegativity difference between the oxygen atom and the remaining H is higher than the electronegativity difference between O and C. However, the difficulty in correctly positioning H_2O in any acidity scale is well-known. The pK_as measured in water give the reverse order for the alcohols, that is the usual order when the solvent effect is considered.

The evaluation of the nucleofugacity causes problems which are

Table 2. Calculated Values of Nucleofugacity in Comparison with Measured Data.

Molecule	SMANIA	Exp.	ACCDEN[a]
C_2H_5-N_2^+	13903	1[b]	4375
CH_3CO-I	6944		
CH_3CO-Br	3012		
CH_3CO-Cl	2258		
CH_3CO-F	346		
C_2H_5-I	1508	2[b]	
C_2H_5-Br	1395	3[b]	
C_2H_5-Cl	986	4[b]	
C_2H_5-F	26	7[b]	
C_2H_5-SCH_3	1100		
C_2H_5-$S^+(CH_3)_2$	2720	6[b]	
C_2H_5-OCH_3	20	9[b]	
C_2H_5-$O^+(CH_3)_2$	570	5[b]	
C_2H_5-OH	3	8[b]	
C_2H_5-$N(CH_3)_2$	12	10[b]	
NCC_2H_4-$N^+(Ph)(CH_3)_2$	540	10.7[c]	1976
NCC_2H_4-SO_2Ph	804	9.6[c]	2713
NCC_2H_4-OPh	447	8.2[c]	3071
NCC_2H_4-$OP(OC_2H_5)_2$	2130	8.0[c]	1276
NCC_2H_4-SPh	10320	7.9[c]	1542
NCC_2H_4-SOPh	4328	(7.1)[d]	2339
NCC_2H_4-OCH_3	131	6.3[c]	1952
NCC_2H_4-$C(NO_2)(CH_3)_2$	8170	(2.6)[d]	2947
NCC_2H_4-OTs	4043		4454
NCC_2H_4-$OCOCH_3$	1690		3327

[a]Acidity levels calculated by the routine ACCDEN. [b]Qualitative order from ref. 35.
[c]Rank values from ref. 13. [d]Values from ref. 13 in which the WG is different from CN.

completely similar to the previous ones, even if in this case there are even less experiences and proposals. Having accepted the fact that the reactions of nucleophilic substitution are influenced by many factors, among which nucleofugacity is not the principal one, and that the nucleofugacity is not easily measured separate from the specific reaction, the availability of a qualitative methodology that permits the ordering of the different leaving groups could suffice.

In Table 2 some examples of nucleofugacity calculation are reported. The first group is composed of usual organic LGs; if the charged groups are eliminated, the ordering is more than acceptable. The LGs containing charged atoms are, to the contrary, very underestimated; this result is however correct if it is considered that the calculation is made on the starting molecule where the stability of the LG, after its departure, cannot be predicted. Therefore considering the series: N_2^+, I, Br, Cl, F, OCH_3, OH; the calculated nucleofugacities are: 13903, 1508, 1395, 986, 26, 20, 3 in good agreement with the only available data[12] that give the qualitative order: 1, 2, 3, 4, 7, 9, 8.

Table 3. Interference values for Boromycin precursor synthesis.

Molecule	Atom	Sign	Activation
1	4	+	2007
	4	-	1033
	8	-	2195
2	1	-	1260
	6	-	3102
3	12	+	2943
	8	-	2193
	11	-	1383
4	3	+	1679
	1	-	2198
	2	-	1686
5	2	+	62
	3	+	62
	5	-	2232
6	1	-	1333
	3	-	1133
	4	-	2195
7	8	+	3006
	17	+	2829
	1	-	1361
	13	-	2231
	15	-	1631
	16	-	2199
8	2	+	1481
	8	-	2195
	9	-	2195

Molecule and Atom numbers correspond to Figure 1. Sign is the qualitative interference type as grasped from the experiment. Activation is either the positive (nucleofugacity or electrophilicity) or the negative (acidity or nucleophilicity) interference calculated by the procedure.

The second group of reported molecules has been analysed so as to compare the results with some more experimental literature data[13]. These measured the nucleofugacity in elimination reactions. The order obtained using only the SMANIA routine is clearly in contrast with the experimental data. This is due to the mechanism of the experimental reaction used that depends not only on the intrinsic nucleofugacity level but mainly on the real reaction mechanism (E2, E1cbR, E1cbI), making a direct comparison very difficult. If, for example, besides the quantity value calculated by SMANIA, the stabilization of the forming negative charge on a neutral LG is considered (i.e. if its stability as a base is considered), a different order is obtained, which is rather more similar to the experimental values.[36] Moreover, considering the series: OPh, $OP(OC_2H_5)_2$, OCH_3, OTs, $OCOCH_3$; the calculated values are: 447, 2130, 131, 4043, 1690; whereas the

measured ones are: 8.2, 8.0, 6.3, n.a., n.a. It is interesting that, in contrast with the reported data, the calculation gives results in the usually expected qualitative order, i.e. OTs > $OP(OC_2H_5)_2$ > $OCOCH_3$ > OPh > OCH_3.

The second part of the discussion regards the interference analysis in medium complexity molecules during the synthesis planning. Two cases will be proposed: 1) the first shows the work of the procedure applied to a synthesis taken from the literature[37] (synthesis of an intermediate of Boromycin, TGT1); 2) the second regards one of the synthesis plans for a structure derived from Strophanthidin (TGT2) proposed by Lilith with particular attention to the interferences.

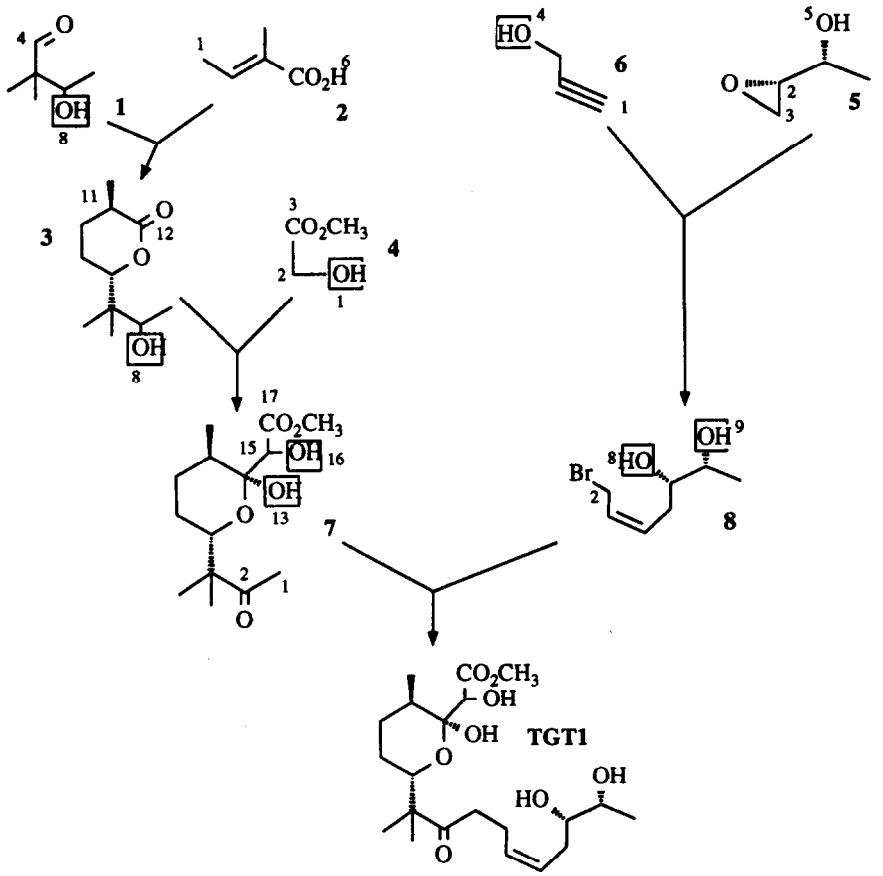


Fig. 1. Scheme of the Boromycin precursor synthesis. The boxed groups have been experimentally protected

In Figure 1 is briefly sketched the experimental synthesis of TGT1. The functional groups that have been experimentally protected are highlighted. In Table 3 the interference values are reported, obtained by the procedure applied to the different unprotected precursors. We will analyze only C-C bond formations.

The first step is the bond formation obtained by nucleophilic addition of a carbanion (atom 1) to an aldehyde group (atom 4). The determined interferences are related only to the formation of the carbanion in gamma position with respect to the acid; the potential SGIs are: the alcoholic oxygen (atom 8) and the

aldehydic proton on **1**, and the acid hydroxyl (atom 6) on **2**. The procedure signals only one SGI for the acidity of the aldehydic proton. Experimentally the aldehyde does not interfere[38], while the (very reactive) acid is protected by its transformation into an anion and the alcohol is etherified (this last group could be as well protected by its transformation into the corresponding anion).

Also the second step is a condensation reaction. The lactone (atom 12) is subjected to the nucleophilic addition without any problem, as it is the formation of the anion on molecule **4** (atom 2), where the present alcoholic group is much more reactive than the carbon and therefore it is not an SGI. Experimentally the OHs on **3** and **4** are protected (see the previous comment).

The preparation of **8** is done in a single step by the addition of an acetylene atom (atom 1) on an epoxide atom (atom 3). The SGIs regard only the possibility of adding to the other epoxide side (atom 2), a fact which is usually avoided for geometrical reasons. Also here the OHs are much more acid than the acetylene (therefore they are not SGIs), but they are experimentally protected.

The last step is a nucleophilic substitution reaction. The bromine in allylic position (atom 2) is a good leaving group and does not give SGIs. The anion formation on **7** (atom 1), on the other hand, can give two kinds of SGIs: a) with competing anions (the alpha-ester proton (atom 15)); b) with potentially electrophilic groups that can give selfcondensations (the activating ketone (atom 2) and the ester (atom 17)). The control of the reaction condition can eliminate both the interferences and, in fact, experimentally the anion preparation does not give problems. On the other hand, the OH groups are experimentally protected, even if their interference could be avoided as mentioned previously.

This first example has given results in agreement with the approach philosophy. The signalled SGIs only regard the structural problems; in the case of anion formation the much more reactive groups are correctly identified but they are not SGIs.

The second example is the two step synthesis of TGT2, as reported in Figure 2 with the interferences reported in Table 4. The synthesis plan, in this case, depends on the previous Lilith routines and therefore the SGIs are influenced.

The level of structural indetermination coming from the analysis (generic LGs and WGs) does not make it possible to establish with certainty whether the SGIs are verified experimentally; however the points with a probable manifold reactivity are signalled and thus they give the information needed to continue the analysis. Let us discuss only some concepts.

During the interaction of molecules **3** (atom 19) and **4** (atom 18) the polarities chosen by the program are both negative; moreover the experimental anion formation in the desired position could cause problems to both structures (many positions more acidic in both the molecules that are not SGIs). But the experienced chemist can see that a substitution on structure **4** is however possible, whereas the dianion formation on **3** could give the desired result; therefore, even if the SGI evaluation does not give the solution, it suggests that the anion on atom 19 (Activation = 1381) will be more easy to prepare, thus indicating the necessity to umpole the reactivity of atom 18 (Activation = 569), a solution of the problem at an even higher level.

The reaction between **1** and **2** seems to be full of problems related to the simultaneous presence of many undefined situations (3 LGs on atoms 4, 9, 12) and of many carbonyls (atoms 6, 9, 25, 35). However by analyzing the bond formation order[8], we can note that, applying the needed structural changes (structures **1'** and **2'**, that have been suggested by us) and the route shown in Figure 3, it is possible to envisage the omission of one LG on atom 4 (using two FGI steps on an already present group) and the use of intramolecular protection to manage alternative reactivities. The fundamental aspect is however the signalling of the SGIs even if the real precursor is not available, thus permitting its better planning.

As far as we know, the only one CAOS program that takes real care of interfering functionalities is LHASA[1,2]. It has two different location mechanisms. 1) The first is inside the transform database. It can be compared with our definition of SGIs; in fact, the functional groups that can give interference because their reactivity is similar to that of the group partecipating in the reaction are explictly cited in each transform. The difference with the present approach can be found in two factors: the first is the necessity

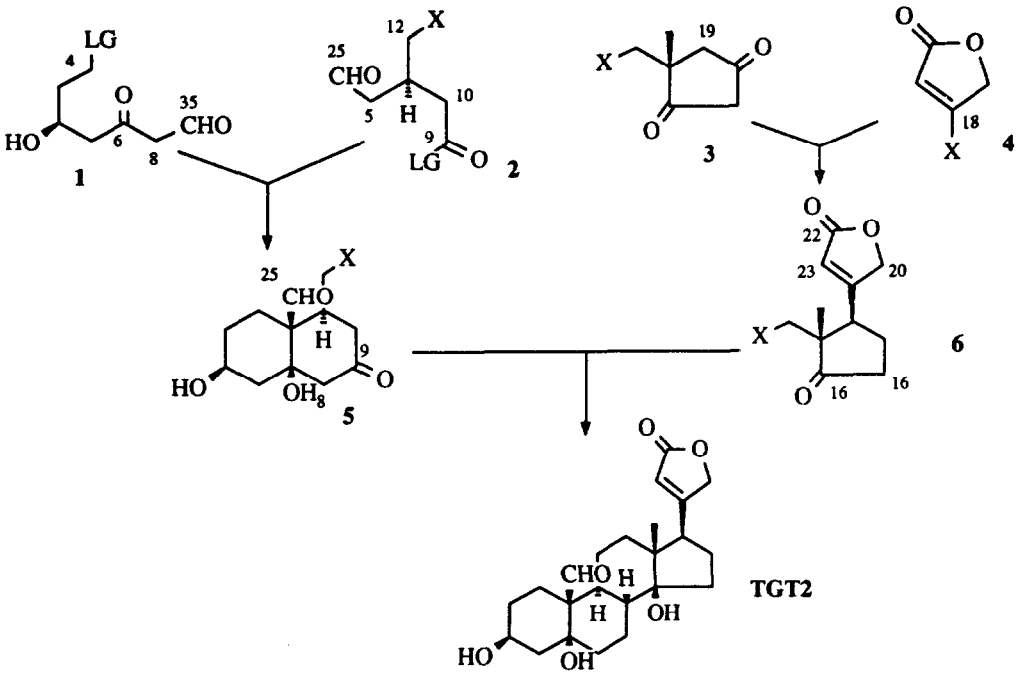


Fig. 2. Synthesis scheme of Strophanthidin

for the transform writer to signal all the possible interfering FGs; the second is the need to use, in any case, the FG definition, i.e. the LHASA approach cannot identify interferences outside its domain. On the other hand, the control operated by the transform writer and the exact definition of what can interfere (i.e. the FGs) furnish results that are very appealing to the chemist. 2) The second considers the interference of the FGs with the prototype reaction conditions suggested by the program. This analysis is based on the previously extensive definition of the interference levels between each FG and each reaction condition. This kind of interference can be compared with the GEI and is not directly comparable with our SGIs.

It must be emphasized that this second interference mechanism, together with a very clever approach to the suggestion of the protective groups needed over multistep syntheses, is actually the most feasible and reliable approach to the GEI analysis.

Table 4. Interference values for Strophanthidin synthesis.

Molecule	Bond break	Atom	Sign	Activation
1	5-6	6	+	3011
		35	+	1991
2	5-6	9	+	6908
		25	+	2012
		5	-	1366
		10	-	1360
		25	-	1036
1	4-5	4	+	1508
		8	-	1370
		35	-	1025
2	4-5	9	+	6908
		12	+	1511
		10	-	1362
		25	-	1033
1	9-8	8	-	1370
		35	-	1025
2	9-8	9	+	6984
		10	-	1362
		25	-	1033
3	18-19	19	-	1381
4	18-19	18	-	569
5	12-13		U[a]	
6	12-13		U[a]	
5	10-15	9	+	3028
		25	+	1971
		8	-	1394
		25	-	1044
6	10-15	15	+	2990
		22	+	1776
		10	-	1375
		16	-	1356
		18	-	1283
		20	-	1610
		23	-	1675

[a] U means Undefined polarity. In this case the program does not examine interferences.
The molecule and atom numbers refer to Figure 2. Bond breaks are given in order of synthetic bond formation. Sign refers to the atom polarity calculated by Lilith. Activation is either the positive (nucleofugacity or electrophilicity) or the negative (acidity or nucleophilicity) interference calculated by the procedure.

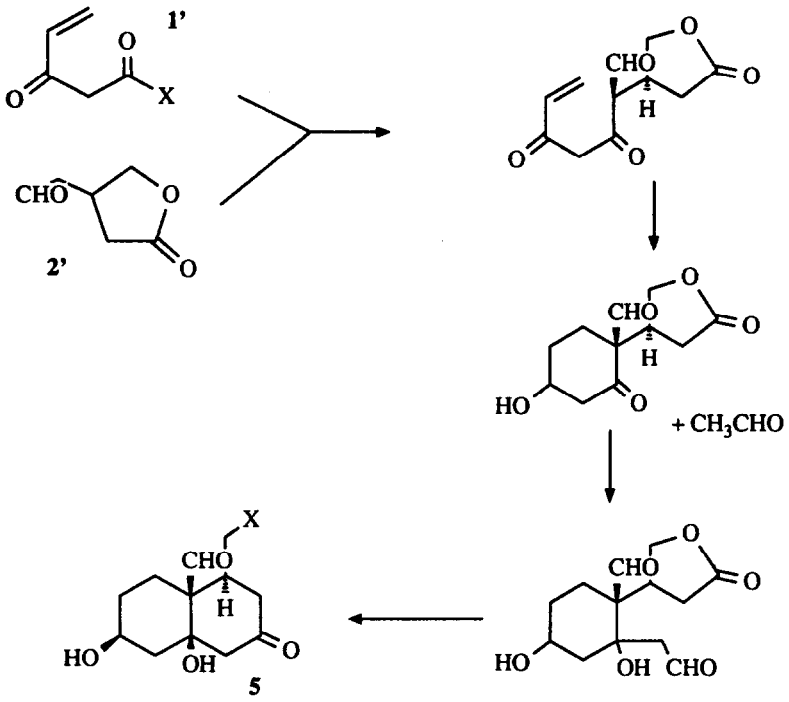


Fig. 3. Synthesis of intermediate 5 of Figure 2

CONCLUSIONS

The evaluation of the possible interferences in synthesis planning is a fundamental part of any kind of theoretical logical approach. A clear separation between SGIs and GEIs allows for the best use of available knowledge, dividing what is related to the molecular structure from what is manageable during the reaction condition planning. The power, given by a theoretical modular model for performing analyses with an increasing level of specification and of repeating some parts of it at different moments, represents a necessary condition to avoid the excessive specificity of the model, removing generality and simplicity while maintaining its efficiency and reliability.

At the present definition level of the introduced approach, the obtained results are in agreement with the main philosophy and, even if they remain very far from direct experimental applicability, permit a better understanding of the synthetic problem and give the information necessary for proceeding with the synthetic analysis.

Acknowledgement. This work has been partially financed by Tecnofarmaci S.p.A., Consiglio Nazionale delle Ricerche e Ministero dell'Universita' e della Ricerca Scientifica e Tecnologica.

APPENDIX 1. LIST OF ABBREVIATIONS USED IN THE TEXT.

ACCDEN	Name of the procedure which calculates acidity (ACidity from Charge, charge Delocalization, and ElectroNegativity).
ALPHAH	ALPHAHardness. The local hardness of atoms.
ANION	The atom which will take the negative polarity on bond breaking.
ATTACK	The atom which will sustain an addition reaction.
BO	The distance between the reactive atom and the atom which participates to the delocalization of the charge measured in conjugations.
BORD	The Bond ORDer of the broken bond.
CENTRE	The atom which is the CENTRE of the considered group of atoms.
CONJP	The contribution to charge stabilization given by each functional group which is electron withdrawing and which is directly attached to the reactive atom
CONJPL	The total contribution to charge stabilization given by the functional groups which are electron withdrawing and which are directly attached to the reactive atom.
CUTEN	The scaling factor used in the calculation of the charge delocalization contribution to the acidity.
DECAY	The damping factor used in the calculation of the charge delocalization contribution to the acidity
DEFINED	The level of definition of the polarity of the atoms spanning the broken bond (DEFINED = no activation required).
DELTAH	The difference (DELTA) in AlphaHardness between two atoms.
FG	A generic Functional Group.
G	The number of functional groups that inductively contribute to the acidity.
GEI	Group Environment Interference. The interference of a functional group with the reaction conditions.
INTERF	The procedure which calculates the group interference.
LEAVING	The atom which will sustain a substitution reaction.
LG	A generic LEAVING group.
MKPREC	The procedure which generates the precursors (MaKe PRECursor).
NG	A generic group which contributes (positively or negatively) to the nucleofugacity of a group.
PCON	A constant used in the acidity calculation (delocalization contribution).
PDEN	A constant scaling factor used in the acidity calculation (inductive contribution).
P(G)	A variable scaling factor used in the acidity calculation (inductive contribution).
PQ	A scaling factor used in the acidity calculation (total contribution).
PSUB	The electronegativity of sp^3 carbons.
P1	A scaling factor used in the acidity calculation (delocalization contribution).
SGI	Similar Group Interference. The interference between functional groups.
SIZE(RING)	The SIZE of a RING.
SMANIA	The procedure which calculates the nucleofugacity of a group.
SUGGESTED	The level of definition of the polarity of the atoms spanning the broken bond (SUGGESTED = activation required).
TGT1,2	The molecule which is the TarGeT of the synthesis planning.
UNDEFINED	The level of definition of the polarity of the atoms spanning the broken bond (UNDEFINED = activation required and polarity to be decided).
WG	A generic electron Withdrawing Group.
χ	The atomic electronegativity.

REFERENCES AND NOTES

1. Corey, E.J.; Orf, H.W., Pensak, D.A. *J. Am. Chem. Soc.* **1976**, *98*, 210-221.
2. Corey, E.J.; Long, A.K.; Greene, T.W.; Miller, J.W. *J. Org. Chem.* **1985**, *50*, 1920-1927.
3. Hendrickson, J.B.; Miller, T.M. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 403-408, and refs. cited therein.
4. Gasteiger, J.; Ihlenfeldt, W.D.; Rose, P.; Wanke, R. *Anal. Chim. Acta* **1990**, *235*, 65-75, and refs. cited therein.
5. For a general review on Computer Aided Organic Synthesis: Vernin, G.; Chanon, M. *Computer Aids to Chemistry*; Ellis Horwood Limited: Chichester, 1986; pp. 19-89.
6. Baumer, L; Sala, G.; Sello, G. *Anal. Chim. Acta* **1990**, *235*, 209-214.
7. Baumer, L.; Sala, G.; Sello, G. *Tetrahedron* **1988**, *44*, 1195-1206.
8. Baumer, L.; Sala, G.; Sello, G. *Tetrahedron* **1989**, *45*, 2665-2676.
9. Baumer, L.; Sala, G.; Sello, G. *J. Am. Chem. Soc.* **1991**, *113*, 2494-2500.
10. Therefore only the SGIs involving the atoms potentially positive or negative will be examined and discussed; on the other hand, it is clear that also transformations governed by a non strictly heterolytic mechanism, but with a different electronic distribution on the reacting sites, are implicitly considered.
11. As will be seen in the SMANIA description part, the ALPHAH determines the goodness of a leaving group and it is therefore correlated to the addition-substitution reactions.
12. Alexander, R.; Ko, E.C.F.; Parker, A.J.; Broxton, T.J. *J. Am. Chem. Soc.* **1968**, *90*, 5049-5069.
13. Stirling, C.J.M. *Acc. Chem. Res.* **1979**, *12*, 198-203.
14. Salatin, T.D.; Jorgensen, W.L. *J. Org. Chem.* **1980**, *45*, 2043-2051.
15. Lewis, E.S. *Bull. Soc. Chim. Fr. II* **1988**, *57*, 259-262.
16. Shorter, J.: Nucleophilic Aliphatic Substitution. In *Organic Reaction Mechanisms 1988*; Knipe, A.C.; Watts, W.E. Eds; John Wiley and Sons Ltd.: New York, 1990; pp. 357-385.
17. Shi, Z.; Boyd, R.J. *J. Am. Chem. Soc.* **1991**, *113*, 2434-2439.
18. Baumer, L.; Sala, G.; Sello, G. *J. Comput. Chem.* **1990**, *11*, 694-699.
19. The term "alpha sphere" is representative of the set of atoms directly connected to the central atom.
20. Sello, G. *J. Am. Chem. Soc.* **1992**, *114*, 3306-3311.
21. Pearson, R.G.; Dillon, R. L. *J. Am. Chem. Soc.* **1953**, *75*, 2439-2443.
22. Clark, J.; Perrin, D.D. *Quart. Rev.* **1964**, *18*, 295-320.
23. Bohme, D.K.; Lee-Ruff, E.; Young, L.B. *J. Am. Chem. Soc.* **1972**, *94*, 5153-5159.
24. Bordwell, F.G.; Algrim, D. *J. Org. Chem.* **1976**, *41*, 2507-2508.
25. Bordwell, F.G. *Pure Appl. Chem.* **1977**, *49*, 963-968.
26. Bordwell, F.G.; Cornforth, F.J. *J. Org. Chem.* **1978**, *43*, 1763-1768.
27. Bordwell, F.G.; Bartmess J.E. *J. Org. Chem.* **1978**, *43*, 3101-3107.
28. Bordwell, F.G.; Fried, H.E. *J. Org. Chem.* **1981**, *46*, 4327-4331.
29. Gushurst, A.J.; Jorgensen, W.L. *J. Org. Chem.* **1986**, *51*, 3513-3522.
30. Taft, R.W.; Bordwell, F.G. *Acc. Chem. Res.* **1988**, *21*, 463-469.
31. Jorgensen, W.L.; Briggs, J.M. *J. Am. Chem. Soc.* **1989**, *111*, 4190-4197.
32. Bordwell, F.G.; Fried, H.E.; Hughes, D.L.; Lynch, T.; Satish, A.V.; Whang, Y.E. *J. Org. Chem.* **1990**, *55*, 3330-3336.
33. March, J. *Advanced Organic Chemistry*, 3rd ed.; John Wiley and Sons Inc.:

New York, 1985; pp. 218-225.
34. Lowry, T.H.; Schueller Richardson, K. *Mechanism and Theory in Organic Chemistry*; Harper and Row: New York, 1976; pp. 124-163.
35. March, J. *Advanced Organic Chemistry*, 3rd ed.; John Wiley and Sons Inc.: New York, 1985; pp. 315.
36. Therefore a transition state in an advanced state of bond breaking is considered.
37. Wierenga, W.: The Total Synthesis of Ionophores. In *The Total Synthesis of Natural Products*; ApSimon, J. Ed.; John Wiley and Sons Inc.: New York, 1981; pp. 296-299.
38. The acidity of aldehydic protons is well-known. Chandrasekhar, J.; Andrade, J.G.; Schleyer, P. von R. *J. Am. Chem. Soc.* **1981**, *103*, 5609-5612.