J. Am. Chem. Soc. **1991**, *113*, 2494–2500

Computer-Assisted Organic Synthesis Planning: Effective Bond Polarity as a Guideline to Reactivity

Luca Baumer, Giordano Sala, and Guido Sello*

Contribution from the Dipartimento di Chimica Organica e Industriale, Università degli Studi di Milano, Via Venezian 21, I-20133 Milano, Italy. *Received October 16, 1989*

Abstract: A first section involved in reactivity prediction has been introduced in a program for computer-aided organic synthesis planning. The concepts of native polarity, induced polarity, and effective polarity have been used to determine the best electronic state for the retrosynthetic breakage of strategic bonds. The characterization obtained is used to suggest the most feasible reactive behavior for the atoms spanning the strategic bond. The procedure and the atomic and molecular physical properties used in the determination of the polarities are described. The logical component is introduced to extend the number of bonds meaningfully treated by the procedure. Results obtained show the low number of bonds that remain unresolved by the semirecursive program logic.

Introduction

LILITH is a program for organic (retro)synthesis planning, which is currently developed in our department.

In previous papers some features of LILITH's chemical structure perception[1] were described. The basic principle of its strategy[1a] is the convergence. The identification of the molecular center is fundamental to define the best convergence of the synthetic scheme, and thus a special metrics for "measuring" molecules has been built. It is based on the definition of the distance between two atoms as a "complexity distance", i.e., "the distance between two atoms A and B in a molecule is equal to the sum of the atomic complexities of the atoms found going from A to B through the bonds connecting A and B". A distance can be calculated for each bond path connecting A and B; the smallest one among them is defined as the minimum distance between A and B. Atomic complexity is calculated by considering both the number of connections of the atom and its stereochemical state (isolated stereocenter, stereocenters α or β to one another, either on chains or on rings). The molecular complexity center is identified by the set of atoms positioned at half the maximum of all the shortest distances. Using this set of atoms, it is possible to define sets of strategic bonds that represent the synthetic solutions proposed by the program.

Ordering affects the solution space dimension. The solution space is ordered not only by arranging the solutions[1a] but also by putting the bond-forming sequence in order for each solution.[1b] The best order in bond-forming sequences is strictly related to convergence and complexity: bonds (in synthetic direction) must be formed in the order that best reduces the "distance" between the precursors and the target molecule. (Here, the term distance is used only in a qualitative fashion and represents the effect obtained by joining the two precursor molecules, thus reducing their physical distance.) This is realized by first forming the bond that minimizes the maximum complexity distance in a molecule. (The best order of bond-forming sequences will be used in discussing the examples.)

In this paper the description of the reactivity section of LILITH, i.e., those routines which must decide the exact nature of the precursors, the kind of reactions to be performed, the activating groups needed, and the possible interferences present, is begun.

Addressing Reactivity

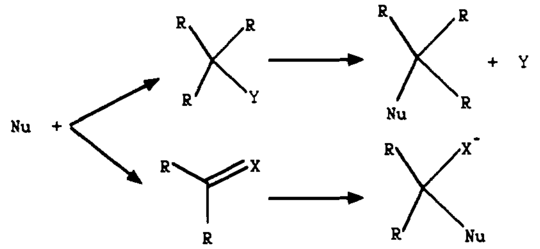
No reaction database is built inside LILITH. This means that the program output makes no explicit reference to any particular reaction. Reactivity is considered in a general way as a set of quantified characteristics associated with the various atoms and regions of a molecule. The word "region" is used instead of "groups" to avoid misunderstandings: no matching or explicit recognition of functional groups is used by LILITH.

Scheme I. Generalized Heteropolar Reaction[a]

$$\mathrm{Nu1} + \mathrm{Pos} - \mathrm{Neg1} \longrightarrow \mathrm{Neg2} - \mathrm{Pos} + \mathrm{Nu2}$$

[a] Nu1 and Nu2 represent generalized nucleophiles, and Pos is the positive and Neg the negative end of the bond.

Scheme II. Addition or Substitution: Alternatives for Generalized Reactions



Inside most subroutines and subroutine blocks of LILITH the IAIA[2] (initial approximation/increasing accuracy) approach is used to get a deeper insight into the synthetic problem as it is processed.

Thus, the reactions of primary interest for LILITH at this processing level are only those devoted to the construction of the structural skeleton of the target. Modifications of the electronic aspect of a particular region (via functional group transforms) will be considered only at a deeper analysis level, after the first level of reactivity interference evaluation. In addition, only heterolytic retrosynthetic transforms are allowed[3] together with ionic models for cycloadditions and concerted reactions.

In the most general treatment,[7] all heteropolar construction reactions are represented as the interaction between the excess

---

(1) (a) Baumer, L.; Sala, G.; Sello, G. *Tetrahedron* **1988**, *44*, 1195–1206. (b) Baumer, L.; Sala, G.; Sello, G. *Tetrahedron* **1989**, *45*, 2665–2676. (c) Baumer, L.; Sala, G.; Sello, G. *Tetrahedron Comput. Methodol.* **1989**, *2*, 37–46; **1989**, *2*, 93–103; **1989**, *2*, 105–118. (d) Baumer, L.; Sala, G.; Sello, G. *Gazz. Chim. Ital.* **1988**, *118*, 745–747.

(2) The IAIA (initial approximation/increasing accuracy) principle represents a particular formulation of a well-known logic principle that suggests reducing the number of possible solutions, increasing the reduction accuracy while decreasing the total number. At the beginning of an OSP activity, the number of proposals could be high but the level of known features about the plan is low. Thus, the rules initially used to cut the solution number down are approximated. As the quantity of valued variables increases and that of residual solutions decreases, the rules must become more and more accurate. This principle operates in the program controlling the program flow, memorizing the logic operations made, and allowing a walk-back where it is required.

(3) Some formalisms, though not completely successful, to "regard all chemical actions as ultimately ionics" were proposed many years ago; see, e.g., refs 4–6.

(4) Lowry, T. M. *Philos. Mag.* **1923**, *46*, 964–976; *Chem. Abstr.* **1923**, *18*, 621.

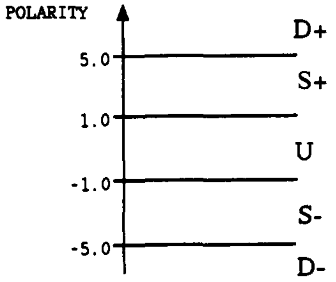
(5) Hartung, E. J. *Chem. Eng.* Mining *Rev.* **1924**, *17*, 34–38; *Chem. Abstr.* **1924**, *19*, 1643.

(6) Prevost, C.; Kirrmann, A. *Bull. Soc. Chim.* **1931**, *49*, 194–243; *Chem. Abstr.* **1931**, *25*, 3309.

(7) Hendrickson, J. B. *Acc. Chem. Res.* **1986**, *19*, 274, and references cited therein.

0002-7863/91/1513-2494$02.50/0

Scheme III. Schematic Representation of the Subdivision of the Polarity Definition Space



of electron density on the negative reacting atom and the (partial) positive charge on the partner atom, as exemplified in Scheme I, where a generalized heteropolar reaction is shown.

In practice,[8] LILITH's approach splits this scheme into the two possibilities in which the displaced electrons are a π bond (addition to unsaturated systems) and a σ bond (substitution). The two alternatives are sketched in Scheme II.

Polarity

In all of the text of this paper, a wide and liberal use of the term polarity[9] will be made. As used here, it is considered as a quantitative characteristic of a (strategic) bond in a target molecule, from which the best functionalization of the precursors can be inferred. Thus, polarity will be the balanced sum of different electronic properties of the atoms involved in the bond and of their environment, together with terms derived from the topological features of the bond being resolved. Quoting Smyth,[10] we can regard polarity "not as physical fact, but [merely] as a pragmatic representation of chemical behavior".

Three different quantities can be specified. Native polarity (NP) is determined by electronic factors: residual atomic charges, electronegativity, conjugation. Induced polarity (IP) is logically derived from native polarities of neighboring atoms and bonds. Effective polarity (EP) is the sum of these two and is the entity used to identify the characteristics required in the reacting molecules to obtain the desired bond.

Polarities (native as well as induced) are calculated as atomic properties, and so the bond polarity is taken as the difference between the polarities of the two atoms joined by that bond. It must be emphasized that the atomic polarities strictly depend on the strategic bond considered; thus, it may happen that the same atom gets different polarities if involved in different bond breaks (e.g., see Figure 1, morphine, where atom 12 has different polarities in bonds 12–15 and 5–12).

Polarity values[11] are calculated in a numerical continuum, but when they are analyzed, they are often considered as "quantized"; i.e., the polarity axis is divided by four threshholds into five regions, symmetrical with respect of the zero point. Scheme III is a representation of the polarity axis divided into these regions.

An atom whose native polarity falls in the "defined positive" (D+) region is a strong electrophile, "suggested positive" (S+) atoms are electrophiles requiring activation, "suggested negative" (S–) atoms are weak nucleophiles, and "defined negative" (D–) atoms are either strong nucleophiles or atoms that can easily generate an anion. "Undefined" (U) polarities indicate that no particular features emerged from the target analysis and that the two possible opposite polarity combinations for two atoms A and B, i.e., A+/B– or A–/B+, are equally feasible, since they both demand a similar level of activation; this situation is characteristic of those parts of the target molecule whose skeleton contains only saturated carbon atoms.

Scheme IV. Equations Used in Native Polarity Definition[a]

```
                                  <---- K1 = -5.0 for case (a) atoms
1)  NP = NP + K1*ΔX              <
                                  <---- K1 = +7.0 for case (c) atoms

2)  NP = NP + K2*ΔX/BORD                K2 = +10.0 for case (b) atoms

3)  NP = NP + K3*ΔX/DECAY               DECAY = BORD^2 * 1.35, K3 = +10

                                  <---- K4 = -2.0 for case (a) atoms
4)  NP = NP + K4*BORD            <
                                  <---- K4 = -1.0 for the other atoms

                                  <---- K5 = -10.0 for the less positive
                                  <                    atom
5)  NP = NP + K5*(Qother-Qcurr)  <
                                  <---- K5 = +10.0 for the more positive
                                  <                    atom
```

[a] *K*'s are the chosen constants, *X*'s are calculated atom electronegativities, BORD's are calculated bond orders (BORD is a negative quantity for conjugated atoms), DECAY is the decaying factor, and *Q*'s are calculated atomic charges.

The Algorithm

LILITH's module DEFSIG will now be described which calculates NP and IP, evaluates the results, and assigns to the solutions proposed by the strategy block a new ranking according to the estimated ease of formation for the bonds of each solution.

Only polarity calculation will be dealt with, neglecting here all the operational subroutines devoted to handling the computer matrix representation of the target.

In the following the work done by the routine on the molecule of lysergic acid will be considered as an example. (The solutions used in the examples come from the sets proposed by LILITH; they are not the best chemical suggestions but they are the more suitable to explain the working of the routine. In the supplementary material other solutions are reported, and there it is possible to find strategic bond sets more similar to those reported in the literature, e.g., in the case of lysergic acid the solutions in which the benzene ring is untouched.)

1. Native Polarities. The most important electronic data used by DEFSIG are the fractional bond orders and the atomic electronegativities and residual charges calculated by the module RESCHA.[1c] Strategic bonds are examined one at a time, following the breaking sequence decided by the previous module of LILITH.[1b]

1.1. The first factor examined is the electronegativity difference (ΔX; Scheme IV, eqs 1 and 2) between the atom (CURR henceforth) whose polarity is to be determined and the atoms in its α sphere[12] (i.e., the atoms directly connected to it). The difference in electronegativity between any two adjacent atoms will be called ΔX [i.e., $\Delta X = X(\alpha) - X(\mathrm{CURR})$]. Any $\Delta X < 0.35$ is considered meaningless inside the selected approximation level and is considered as equal to 0.

ΔX between CURR and each atom inside its α sphere is multiplied by a parameter, and the resulting α–ΔX value is added to CURR's NP. The choice of the parameters distinguishes between three situations: (a) CURR is less electronegative than the α atom ($\Delta X > 0.35$) AND singly bonded to it AND the α atom does not bear any hydrogen atom (e.g., ethers, tertiary amines) (Scheme IV, eq 1, $K1 = -5.0$). (b) $\Delta X > 0.35$, but one of the other two conditions of the preceding point (or both of them) is not satisfied (e.g., alcohols, imines); in this case the bond order becomes part of the parameter. (Scheme IV, eq 2). (c) $\Delta X < -0.35$ (e.g., CURR is an heteroatom bonded to carbon (Scheme IV, eq 1, $K1 = 7.0$).

Only atom 11 of lysergic acid has an α sphere with sufficient electronegativity (the N atom 16 gives a nonzero contribution)

---

(8) This classification derives from a heuristic classification of overall reaction types and is not meant to be a detailed reproduction of any real mechanism.

(9) March, J. *Advanced Organic Chemistry*, 3rd ed.; Wiley-Interscience: New York, 1985.

(10) Rebek, M. *Monatsh. Chem.* **1952**, *83*, 144; *Chem. Abstr.* **1952**, *46*, 5964.

(11) Typical polarity values lie in the range −10/+10, with some exceptional cases scoring up to ±30. These numbers are considered as dimensionless, since they derive from the sum of different physical entities.

(12) In all of the target processing, the strategic bond is considered as already (retrosynthetically) broken. The valence so freed on each of the two separated atoms is left unsaturated, e.g., the alpha-sphere of an sp^3 carbon atom involved in the strategic bond contains only three atoms. DEFSIG's output will be used by the following interference module to decide which atom or group must be used to saturate it.

and thus its NP = 0 + 0 + (2.928 − 2.527)(−5.0) = −2.0 (Figure 1, row 1).

1.2. The second factor is the possibility of charge delocalization. This situation is detected by the program whenever two atoms respectively α and β to CURR are linked with a calculated bond order[1c] (BORD; Scheme IV, eq 2) greater than 1. ΔX between the α and β atoms is evaluated $[X(\alpha) - X(\beta)]$, and all $\Delta X > 0$ are considered. If the α–β multiple bond is part of a longer conjugated π chain or an aromatic system, all following conjugated multiple bonds are also considered. In this case the ΔXs associated with each multiple bond are added together, each one divided by a distance decay factor (DECAY; Scheme IV, eq 3) proportional to the square of the number of intervening double bonds between CURR and the reached atom. The resulting $\beta - \Delta X$ value, multiplied by a parameter to keep it in scale with the previously calculated $\alpha - \Delta X$ values, is added to NP (Scheme IV, eq 3).

Atom 10 of lysergic acid can delocalize a charge onto the double-bond system (the aromatic N atom 9 gives a nonzero contribution on two paths) and its NP = $2(2.649 - 2.964)/(-1)^2(1.35) = -4.7$ (Figure 1, row 2).

1.3. A third structural element is now checked: the vinylic or arylic position of the breakage. A bond between an sp^2 and an sp^3 carbon (unsaturated and saturated atom) is given a preferential polarity with the sp^2 atom more negative than what was computed by the $\alpha - \Delta X$ factor only, especially if it is part of an aromatic ring. A unique simple function of the bond order takes care of both aryl and vinyl situations (Scheme IV, eq 4), and the final multiplicative parameter attributes larger weight to this factor if the first two (points 1.1 and 1.2) resulted in no definite indication (Scheme IV, eq 4, $K4 = -2.0$, case a). The value obtained is added to NP. Atoms 2 and 3 of lysergic acid are in a benzene ring before their connecting bond is broken, and thus they are considered arylic [NP = −2(1.67)]. Atom 2 is no more arylic when subsequently breaking bond 2–1, whereas atom 1 is still arylic [NP of atom 1 = −2(1.33) − 2(1.33) = −5.3; NP of atom 2 in this break is still 0] (Figure 1, row 3).

1.4. The fourth and final factor involved in native polarity is the difference ($Q_{\text{other}} - Q_{\text{curr}}$; Scheme IV, eq 5) in calculated electric charge[1c] between CURR and its partner, the two atoms linked by the strategic bond.

After the usual multiplication by a scaling parameter (Scheme IV, eq 5), this difference is also added to NP, contributing the same absolute value and opposite signs to the two-atom polarity. The NPs of all five atoms of lysergic acid are changed due to residual charges as could be observed in row 4 of Figure 1 [e.g., NP of atom 11 is NP = $-2.0 + (0.021 - (-0.035))10 = -1.4$].

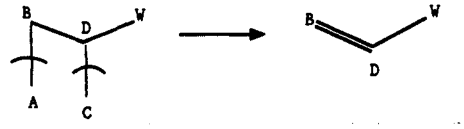
At this point, each atom has been assigned an NP value that classifies it as undefined, suggested (+ or −), or defined (+ or −), giving rise to 25 possible combinations for the two atoms. In fact, it must be stressed that calculating this particular polarity separately for the atoms and not globally for the set atom–bond–atom allows the possibility of having two "poles" with the same sign, which is not the case for customary use and evaluations of bond polarity. (It is possible to find situations where an atom has a polarity value for one strategic bond and is still undefined for a different one.) Some examples are reported in Figure 1.

2. Induced Polarities. Undefined situations are useless for the goal of selecting a reactive hypothesis, so the next task of DEFSIG is to minimize the number of such situations. This is realized by trying to attribute an IP value to atoms with undefined polarities; the induction process is not an electronic effect in this context but an outcome of the following logical analysis of the topology of the molecule.

The subroutine INDUCE considers only two possibilities for each atom, depending on the magnitude of the effective polarity: undefined [i.e., ABS(NP + IP) < 1.0] or set (Scheme III), regardless of the sign and the level of the definition.

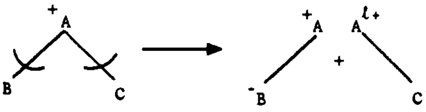
The algorithm is divided into a cascade with three descending (of lesser importance) steps, which need not always be executed. As soon as the undefined condition is removed, INDUCE stops processing that atom. This means that in the following paragraphs it will be always implicitly assumed that the calculations are executed only if CURR is still undefined.

Scheme V. Logical Definition of Vicinal Atom[a]



[a] W is an activating group, D is electronically set, and B is logically set.

Scheme VI. Logical Definition of Geminal Bond[a]



[a] Polarity of A was set in A–B breakage and undefined in A–C breakage.

2.1. The first step is the examination of the partner definition level. If the CURR's partner (the other reacting site, OTHER henceforth) is known to have a preferential polarity (i.e., it is "set"), it is straightforward to assign to CURR an opposite polarity sign. In fact, the reversed choice would give no advantage from the side of the undefined atom, which in any case requires the addition of an activating group, but would involve the umpolung of the already set atom and thus at least two more synthetic steps. OTHER's setting accounts for both polarity magnitudes, i.e. IP(CURR) = −EP(OTHER).

If, on the contrary, OTHER is also undefined, then both IP's are changed in a similar way: IP(CURR) = −NP(OTHER) and IP(OTHER) = −NP(CURR). Since NP is small, this is only very seldom sufficient to set both atoms, but it may happen in those cases in which a high charge difference (see 1.4 above) almost put both atoms into a suggested condition.

This calculation is made for all the bonds of the examined solution, so that now no set atom can be a bond partner with an undefined one; i.e., each bond is either set at both extremes or undefined at both.

Atom 2 of lysergic acid in bond 2–1 has thus an IP = +5.3 that summed to the preceding value gives an EP = +5.2 (Figure 2, row 1).

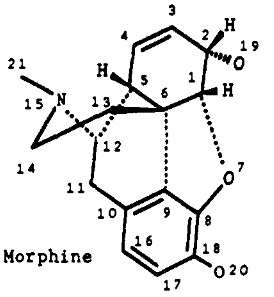
2.2. Next, the relative positions of strategic bonds are examined. If two strategic bonds are vicinal, and one of them is set while the other is not, the cause of the setting is checked and its possible dependence on the conjugation with a multiply bonded group is verified. If this is true, then the undefined atom next to the set bond is set with a polarity opposite to that of the connecting atom: IP(B) = IP(B) − EP(D). This has the effect of preparing the transformation of the two vicinal breakages into a 1,2 addition to a conjugated double bond. This transform is also depicted in Scheme V. If the created double bond is not activated, the transform is not selected because regiospecificity would be missing or is not addressable at this level.

Each undefined bond is checked against the vicinal breakage skeleton, and if any atom previously undefined is now set, the program loops back to 2.1 to keep the two partners in the same condition of setting. This situation is not present in lysergic acid, but considering the molecule of lycopodine it is possible to see that atom 16 is α to atom 7 and IP is 9.7 (Figure 2, row 2).

2.3. A last possibility is that of double affixations[1a] (geminal breakages). If one of the two geminal bonds is set, while the other is not, then the program favors the situation in which the central atom has the same polarity in both the breakages, since the same activating groups[13] needed for one reaction will probably be well suited also for the other one (and in practice the two reactions could be a single synthetic step). The assignment is IP(CURR, second bond) = EP(CURR, first bond) (see Scheme VI). To

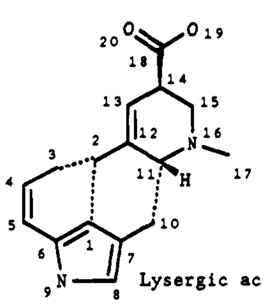
---

[13] If the polarity of the central atom was (electronically, see 1) set without asking for additional activating groups, it would have been set in both bonds in the same step.



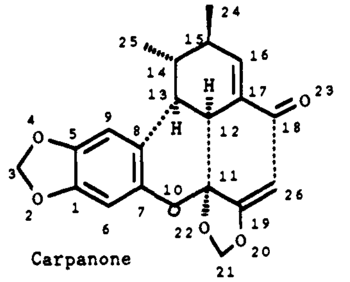
Morphine

BOND	15 - 12		1 - 7		5 - 12		6 - 9	
Δ χ	-5.8			-4.5				
Qdel	-5.8			-4.5	-.01			
ar/vn	-5.8			-4.5	-.01		-6	
Δ Q	-7.2	1.4	3.2	-7.7	-.34	.33	-6.1	-.07
Lod	D-	S+	S+	D-	U	U	D-	U



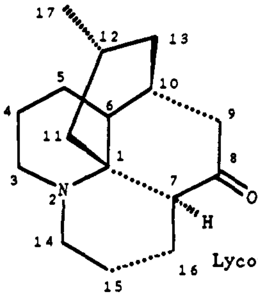
Lysergic acic

BOND	11 - 10		3 - 2		2 - 1	
Δ χ	-2.0					
Qdel	-2.0	-4.7				
ar/vn	-2.0	-4.7	-2.7	-2.7		-5.3
Δ Q	-1.4	-4.2	-2.3	-3.1	-.1	-5.3
Lod	S-	S-	S-	S-	U	D-



Carpanone

BOND	26 - 18		11 - 12		8 - 13	
Δ χ		4.3	-7.3			
Qdel		4.0	-14.			
ar/vn	-4.	4.0	-14.		-6.	
Δ Q	-7.3	7.2	-11.	-2.5	-5.9	.1
LoD	D-	D+	D-	S-	D-	U



Lycopodine

BOND	10 - 9		15 - 16		1 - 7	
Δ χ					-2.	
Qdel		-9.3			-2.	-9.3
ar/vn		-9.3			-2.	-9.3
Δ Q	-.1	-9.2	.	.	-1.6	-9.7
LoD	U	D-	U	U	S-	D-

Figure 1. Example of native polarity calculation. The values on each line are the result of the contributions given by all the variables considered in the preceding lines. The bonds are examined in breaking sequence. The variables have the following meanings: BOND, strategic bonds for the examined solution; Δ*X*, electronegativity difference contribution; Qdel, charge delocalization contribution; ar/vn, aryl or vinyl presence contribution; Δ*Q*, atomic charge difference contribution; LoD, level of definition of the corresponding atom, i.e., defined, suggested, undefined. The values reported represent the current value of the atom polarity in the corresponding strategic bond shown at the top of the figure.

illustrate this point, the molecule of morphine is considered. Atom 12 in bond 5–12 is still undefined (Figure 2, row 2), whereas in bond 15–12 it is suggested positive (S+). IP for atom 12 in bond 5–12 thus becomes equal to 1.4 (Figure 2, row 3).

Again, if an undefined atom is now set, the program loops back to 2.1 and repeats all the logical decision scheme to set as many bonds as possible.

Morphine is again considered. Atom 5 is the partner of atom 12 in bond 5–12, and it is still undefined (Figure 2, row 3) but, due to the last change in EP of atom 12, its EP becomes −1.7 (Figure 2, row 4).

Some examples describing the search for IPs are reported in Figure 2.

3. Conflicting Polarities. The possibility exists that both the strategic bond extremes have the same polarity; i.e., they would preferentially react both as nucleophiles or both as electrophiles. (In the examples in Figure 3 morphine will not be considered as all its atoms have concordant polarities). In this particularly unfortunate situation, an intermediate umpolung passage will be necessarily applied to one of the active sites. The algorithm must now try to decide which one of them will be preferentially chosen.

Three criteria are applied in sequence.

3.1. First, the possibility for both the atoms to be part of a conjugated system is checked. If this is the case, a calculation to decide which of them can best delocalize an electric charge[14] is performed. The residual charges on the neutral fragments are recalculated,[15] and then a unitary electronic charge, positive or negative according to the polarity sign, is placed on the examined atom. The routine RESCHA,[1c] which performs charge distribution, is run again by using the charged structure.

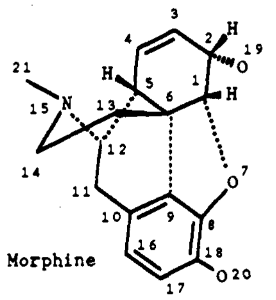
The atomic charge difference between the neutral and the charged fragments is taken as the index of delocalizing capability and, after the usual scaling (by a factor equal to 10.0), is added to NP. The formal unitary charge is then moved from the first to the second atom in the input structure, the calculation is repeated, and the two results are compared. Atoms 10 and 11 of lysergic acid are both (S−) (Figure 2, last row); this last calculation allows a change of EP of atom 11 that is set to 2.0 (S+) (Figure 3, row 1).

---

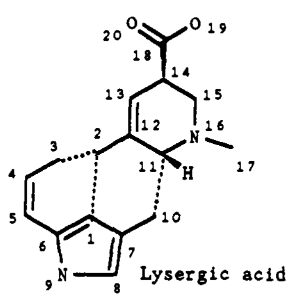
[14] If only one of the atoms is conjugated with a multiple bond, it must be the one which would afford larger delocalization, but it is also the one which would best distribute an opposite charge after an umpolung process and thus the rule cannot be applied.

[15] If one of the broken bonds in the target is part of an aromatic ring, the Kekulē canonical form is chosen which maximizes the number of double bonds in the fragment deriving from ring opening.

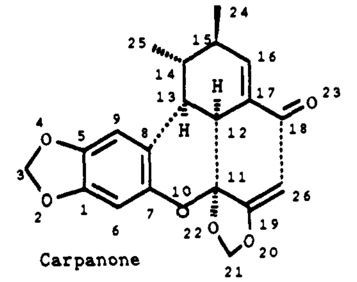
[16] The second and third steps of INDUCE are then repeated to modify the topologically selected settings according to the reversed polarity.



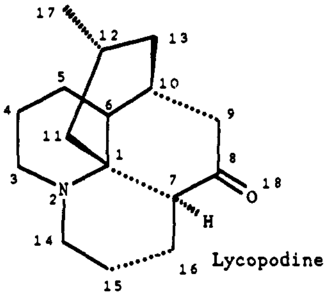
BOND	15 - 12		1 - 7		5 - 12		6 - 9	
PD	-7.2	1.4	3.2	-7.7	-.34	-.33	-6.1	6.1
DB	-7.2	1.4	3.2	-7.7	-.34	-.33	-6.1	6.1
DA	-7.2	1.4	3.2	-7.7	-.34	1.4	-6.1	6.1
Loop	-7.2	1.4	3.2	-7.7	-1.7	1.4	-6.1	6.1
Lod	D-	S+	S+	D-	S-	S+	D-	D+



BOND	11 - 10		3 - 2		2 - 1	
PD	-1.4	-4.2	-2.3	-3.1	5.2	-5.3
DB	-1.4	-4.2	-2.3	-3.1	5.2	-5.3
DA	-1.4	-4.2	-2.3	-3.1	5.2	-5.3
Loop	-1.4	-4.2	-2.3	-3.1	5.2	-5.3
Lod	S-	S-	S-	S-	D+	D-



BOND	26 - 18		11 - 12		8 - 13	
PD	-7.3	7.2	-11.	-2.5	-5.9	5.9
DB	-7.3	7.2	-11.	-2.5	-5.9	5.9
DA	-7.3	7.2	-11.	-2.5	-5.9	5.9
Loop	-7.3	7.2	-11.	-2.5	-5.9	5.9
LoD	D-	D+	D-	S-	D-	D+



BOND	10 - 9		15 - 16		1 - 7	
PD	9.1	-9.2	.	.	-1.6	-9.7
DB	9.1	-9.2	.	9.7	-1.6	-9.7
DA	9.1	-9.2	.	9.7	-1.6	-9.7
Loop	9.1	-9.2	-9.7	9.7	-1.6	-9.7
LoD	D+	D-	D-	D+	S-	D-

Figure 2. Example of induced polarity calculation. The values on each line are the result of the contributions given by all the variables considered in the preceding lines. The bonds are examined in breaking sequence. The variables have the following meanings: BOND, strategic bonds for the examined solution; PD, partner definition effect contribution DB, activated double-bond effect contribution; DA, double-affixation (geminal breakage) effect contribution; Loop, value contributed by the repetition of the induced polarity search; LoD, level of definition of the corresponding atom, i.e., defined, suggested, undefined. The values reported represent the current value of the atom polarity in the corresponding strategic bond shown at the top of the figure.

3.2. The polarity values of the two atoms, possibly corrected by the delocalization factor, must be compared. The value of a variable σ is calculated as a discontinuous function of the polarity difference [δ = EP(CURR) − EP(OTHER)]. If δ is less than 6.0, σ is a predefined constant (5.0). If δ is greater than 6.0, σ will be the quantity necessary to invert the polarity of the less defined [smaller ABS(EP)] atom to a suggested level of opposite sign. Then both EP's are diminished:

$$\mathrm{EP(CURR)} = \mathrm{EP(CURR)} - \sigma \qquad \mathrm{EP(OTHER)} = \mathrm{EP(OTHER)} - \sigma$$

If the atom which had the larger EP value is still set with the same polarity sign it had before, it is considered so strongly defined to justify the umpolung of the partner;[13] otherwise, the EP decrease is undone for both atoms, since the rule cannot assess a preference for any of them.

Lycopodine is considered as an example. Atoms 1 and 7 both have negative polarities, but atom 7 is much better defined; thus, the EP of atom 1 is set to 2.0 (S+) (Figure 3, row 2).

3.3. If the δ rule failed, the program applies a steric rule. Starting from the general consideration that additions to double bonds require less space availability than substitution to atoms with four non-hydrogen atoms attached, the various possible combinations are scanned to find what kind of reaction could be performed to form the examined bond.

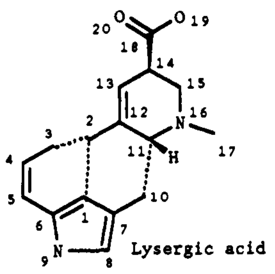
Steric hindrance at the reacting sites of the fragments is evaluated by the subroutine already used to rank and sort the solution set.[1b] If one of the two centers is much more hindered than the other, a polarity inversion is proposed if this allows a substitution reaction to be performed at the least hindered atom or an addition at the most hindered.

Atoms 2 and 3 of lysergic acid have both negative polarities but atom 2 is more hindered and its EP is set to 2.0 (S+) (Figure 3, row 3).

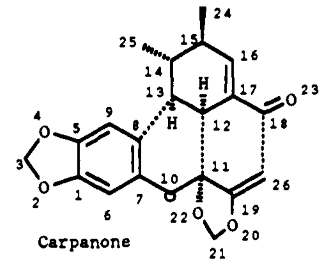
Polarity is also inverted at one atom requiring the insertion of an activating group, being the partner's EP of the defined type, since it is thought possible to select a group of the desired hindrance which is also capable of inverting the low effective polarity of that atom. If a polarity inversion has been devised, the subroutine INDUCE is run again to modify the IP's of the other bonds according to the new situation.

Examples of the treatment of polarity conflicts are reported in Figure 3.

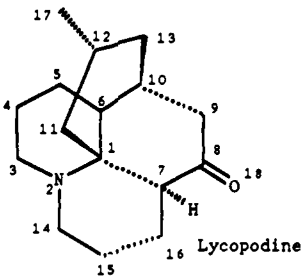
If, on the contrary, no reason is found to select one of the two partners for inversion, the conflicting situation remains unsolved. In this case an additional solution can be added to the solution



BOND	11 - 10		3 - 2		2 - 1	
Qdel	2.0	-4.4	-2.3	-3.1	5.2	-5.3
PW	2.0	-4.4	-2.3	-3.1	5.2	-5.3
Steric	2.0	-4.4	2	-5.	5.2	-5.3
Lod	S+	S-	S+	S-	D+	D-



BOND	26 - 18		11 - 12		8 - 13	
Qdel	-7.3	7.2	-11.	2.	-5.9	5.9
PW	-7.3	7.2	-11.	2.	-5.9	5.9
Steric	-7.3	7.2	-11.	2.	-5.9	5.9
LoD	D-	D+	D-	S+	D-	D+



BOND	10 - 9		15 - 16		1 - 7	
Qdel	9.1	-9.2	-9.7	9.7	-1.6	-9.7
PW	9.1	-9.2	-9.7	9.7	2.	-9.7
Steric	9.1	-9.2	-9.7	9.7	2.	-9.7
LoD	D+	D-	D-	D+	S-	D-

Figure 3. Example of conflicting polarity resolution. The values on each line are the result of the contributions given by all the variables considered in the preceding lines. The bonds are examined in breaking sequence. The variables have the following meanings: BOND, strategic bonds for the examined solution; Qdel, contribution given by the charge delocalization on the ion; PW, polarity weighting contribution; Steric, steric hindrance effect contribution; LoD, level of definition of the corresponding atom, i.e., defined, suggested, undefined. The values reported represent the current value of the atom polarity in the corresponding strategic bond shown at the top of the figure.

set, containing the same bonds as the current one. In one of these solutions, an umpolung is proposed for CURR; in the other one, the umpolung is proposed for the partner.

4. Ranking and Sorting. Eventually, a new evaluation is made for each solution, and another factor is added to the strategic value.[1a] The ideal case is considered the defined positive/defined negative situation. A small worsening parameter (1.11) is used for only suggested cases, and a larger one (1.32) for atoms whose polarity has been inverted; then in decreasing order there are the undefined bonds (1.53) and the unsolved conflicts (1.84–2.36).

The solution set is now sorted again, and output is made starting from the new "best" solution. The last solution(s) can be rejected if their strategic value is very different from the best one, as already done after the previous evaluation sections.[1a,b]

5. Technical Remark. LILITH is entirely written in FORTRAN 77 and is implemented on an IBM 3090 (using a VMHPO/FORTVS compiler) and on a Bull X-20 (using a UNIX/SVS FORTRAN compiler). A single version of the program fulfills the two system specifications.

Discussion

The importance of polarity in reactivity predictions is well established.[4-6,17] The present one is an approximation since the characteristics of the transition state are neglected, and they are the most important for an accurate examination of the reaction course. Some of the values used in the routine DEFSIG to calculate polarity are derived from the target and not from the precursors, but this cannot augment the approximation level since, for a not-yet-known reaction, the product could as well be more similar to the transition state than the reactants. Thus, examining some possible precursors would be useless.

A wider test, made with tens (or hundreds, if many different reaction mechanisms are to be investigated) of precursors and well-defined selected reactions would probably yield more precise results, but it would be completely opposing LILITH's basic philosophy. Instead, the IAIA approach can afford as many reliable predictions by indicating how the reactants and the reaction conditions must be chosen to realize the addressed reactivity.

LILITH's assumption is that complementary reactivity is indicated by differences in polarity of the atoms that will be directly bonded; this can be seen as a corollary of Carre's rule.[18] The topological origin of the induced factor in polarity calculation does not necessitate two different interpretations, since an atom's IP can be seen as the predicted NP of that atom after the functionalization transform.[19] The greatest ease of bond formation will be found in the juxtaposition of two large EP values with opposite signs, i.e., of the largest (atomic) EP algebraic difference, since this situation can be regarded as more "activated".[20]

---

(17) Hellstrom, N. *Sven. Kem. Tidskr.* **1941**, *53*, 405–412; *Chem. Abstr.* **1941**, *36*, 2788.

(18) Carre's rule affirms: 'The reactive differences of directly linked elements may be considered as corresponding to a polarity difference which increases as the electronic strain increases.' Carre' P. *Compt. Rend.* **1936**, *202*, 740–742; *Bull. Soc. Chim. Fr.* **1936**, *5*, 1073–1982.

(19) Rideal, E. K. *Trans. Faraday Soc.* **1923**, 514–517.

(20) Although "activated" is used with a general meaning[23] of "large driving force", a lower activation energy can be supposed for reactions where a large initial difference in charge produces strong electrostatic interactions and/or a large initial difference in electronegativities to be equalized[21,22] makes steeper the reaction diagram.

(21) Sanderson, R. T. *Science* **1955**, *121*, 207–208.

The actual performance and the practical results of this routine heavily depend on a delicate balance of the different parameters, the scaling and weighting factors, and the thresholds, which are used to quantify and mix the electronic and topologic elements involved in the calculations. The problems inherent to the application of these mathematical/statistic techniques to chemical systems were already pointed out during the development of the first applications of artificial intelligence to analytical chemistry. For example, the simplest of DEFSIG's groupings, i.e., set positive/undefined/set negative classification described under section 2, is analogous to the application of a "width" parameter to a threshold to obtain a "dead" zone, i.e., an uncertainty hypervolume in the decisional hyperspace, as described by Jurs and Isenhour.[23] The fivefold division described under section 1.4 can in turn resemble one of Bender and Kowalski's multiclass linear classifiers.[24]

Four differences should be stressed: (1) Polarities are definitely NOT linearly separable. (2) The undefined polarity case is not an uncertainty zone. It asserts a particular electronic environment as well as the other cases; i.e., also by use of perfect parameters and functions, this area would not tend to empty, and so this important criterion of optimization is missing. (3) The variables involved are not always linearly independent from each other, and some of them (e.g., electronegativity) are used more than once, in different points of the algorithm, with different weights; as a consequence, the systematic variation of the parameters one at a time cannot ensure a good optimization level. (4) It is essential to remember that a quantitative theory of reactivity to compare with is not available, and no transform collection to fit the results onto is inserted in LILITH. This prohibits the application of an automatic or semiautomatic feedback approach to optimization with the aid of a training problem set.

The parameters and the thresholds currently used by LILITH have thus been selected by analyzing by hand[25] the output obtained for a test set of 30 complex natural compounds.[26] The results are satisfactory in that no more than 1% of the classifications are changed by a 10% variation of any single parameter or threshold. The number of undefined bonds is kept small (188 cases on 646 calculations, but 94 cases are for molecules without any functionalization), even in poorly functionalized molecules, by the application of IP's. This is also a positive consideration, because undefined and conflicting polarities cause an exponential increase in the number of solutions to be further processed (every solution containing N undefined strategic bonds generates $2^N - 1$ new solutions).

In addition, the polarities calculated by the procedure are in qualitative agreement with those suggested by considering the functionalization of the molecules. It must be emphasized that the polarity calculation is the subject of the present paper, and the evaluation of the results is here limited to the reliability of the bond polarities obtained. None of them contrasts with the polarity that could be suggested by a chemist, and many of them correctly reproduce the polarities suggested by common reaction mechanisms.[26]

It is possible to refer to Figure 2 for the morphine molecule and Figure 3 for the other structures.

In Figure 2, morphine, atom 15 has a DEFINED negative polarity as expected in a nitrogen–carbon bond, atom 9 has a DEFINED negative polarity as expected in an aryl–alkyl bond, and atom 7 has a DEFINED negative polarity as expected in an oxygen–carbon bond. All the other atoms have logically defined polarities.

In Figure 3, lysergic acid, both atoms 2 and 3 had SUGGESTED negative polarities until the application of the steric rule, as expected for allyl carbons, atom 1 has a DEFINED negative polarity as expected in an aryl–alkyl bond, and atom 10 has a SUGGESTED negative polarity as expected in an aryl–alkyl bond. All the other atoms have logically defined polarities.

In Figure 3, carpanone, atom 8 has a DEFINED negative polarity as expected in an aryl–alkyl bond, atom 11 has a DEFINED negative polarity as expected on an allyl carbon α to an oxygen, atom 26 has a DEFINED negative polarity as expected in an aryl–alkyl bond, and atom 18 has a DEFINED positive polarity as expected for a carbonyl carbon. All the other atoms have logically defined polarities.

In Figure 3, lycopodine, both atoms 7 and 9 have DEFINED negative polarities as expected for α carbonyl carbons. All the other atoms have logically defined polarities.

Conclusions

The possibility of predicting the most feasible effective polarity (EP) for a strategic bond breakage has been demonstrated.

LILITH has thus available a tool to predict a first level of reactive behavior for organic molecules, and the chance to deepen the reactivity analysis has come nearer.

The various concepts of polarity used have established a new meaning for the longstanding idea of polarity present in the organic chemical literature, which is very often undefined and inaccurate.

The definitions introduced do not solve the problem of a characterization, in words, of the concept but give a well-defined procedure to obtain a quantitative description of the empirical polar state for a bond in a construction reaction.

Acknowledgment. Partial financial support by Tecnofarmaci S.p.A. and by the Consiglio Nazionale delle Ricerche is gratefully acknowledged. L.B. and G.S. hold fellowships from Tecnofarmaci S.p.A. Special thanks are due to the referees, whose suggestions greatly improved the present paper.

Supplementary Material Available: Complete LILITH output obtained for the 30 compounds cited (20 pages). Ordering information is given on any current masthead page.

---

(22) Gasteiger, J.; Saller, H. *Angew. Chem., Int. Ed. Engl.* **1985**, *24*, 687–689, and references cited therein.

(23) Jurs, P. C. *Anal. Chem.* **1971**, *43*, 1812–1815. Wangen, L. E.; Frew, N. M.; Isenhour, T. L. *Anal. Chem.* **1971**, *43*, 845–849, and references cited therein.

(24) Bender, C. F.; Kowalski, B. R. *Anal. Chem.* **1974**, *46*, 294–296.

(25) The possibility of using some modern statistical techniques (e.g., Wold, S.; Esbensen, K.; Geladi, P. *Chemomet. Intell. Lab. Syst.* **1987**, *2*, 37) to optimize thresholds and parameters can be considered, even if the approach proposed is highly independent from the calculated values, and the result validation made by expert chemists is one of our preferred working style.

(26) The complete output obtained for the 30 compounds cited is available as supplementary material.