New Method for the Calculation of Bond Native Polarity Using Molecular Electronic Energy

LUCA BAUMER and GUIDO SELLO*

Dipartimento di Chimica Organica e Industriale, Università degli Studi di Milano, via Venezian 21, I-20133 Milano, Italy

A procedure for the calculation of atomic and molecular electronic energy is described. The calculated energies of some example molecules are compared with either ab initio or semiempirical results. The calculation of electronic energies is used inside a different procedure to evaluate the native polarities of strategic bonds of organic molecules; the corresponding algorithm is described, and some examples of its application are reported. The results are compared with those obtained by a previously described approach and with usual chemical reactivity. Differences and improvements introduced by the present method are discussed.

INTRODUCTION

A program (Lilith[1]) for computer-assisted organic synthesis design is being developed in our department. It is composed of two main parts: the strategy[1a] that identifies the best synthetic paths of the solution space and the reactivity predictor that ensures both the best use of the functional groups present in the target molecule and the choice of the best reaction conditions.[2]

In the present paper we discuss one of our procedures used by the reactivity predictor. A new method for the calculation of bond native polarities[3] (NPs) (whose meaning is described in the next section) is introduced; this method uses the molecular electronic energy as the main variable to help in choosing one among the alternative NPs. A new fast procedure for the calculation of the molecular electronic energies is also described. The energies calculated for some example molecules show a good correlation with the results obtained using other much more sophisticated approaches (either ab initio or semiempirical calculations). This comparison is easily possible for molecules in their ground state, but it is very difficult for the simulated breaking of strategic bonds, which gives molecules with charged atoms that have free valences. Other experiences (e.g., Gasteiger group's work[4]) evaluate the polar direction of heteropolar bond breakings, but even in this case a direct comparison is difficult. In fact the validation of their and our results is done directly by the experienced chemist on a trial set of molecules, whereas the variables used to quantify the results are very different. Therefore in the discussion on the validity of the calculated NPs, we compare them with the known molecular reactivity.

BACKGROUND

A brief outline of the Lilith program is done in this section.

The main principle of Lilith strategy is the convergence that is ensured by the identification of the molecular center of the target. For this purpose a special metrics for "measuring" molecules is used; it is based on the definition of the "complexity distance" between two atoms, 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 on the path from A to B". Atomic complexity is calculated considering both the substitution level of each atom and its stereochemistry (i.e., isolated stereocenter; stereocenters α or β to one another, either on chains or on rings). A distance can be calculated for each bond path connecting A and B; the shortest distance is the minimum distance between A and B. The pairs of atoms that are at the maximum of the shortest distances are the most

distant atoms; the set of atoms placed at half of the maximum distance from the most distant atoms is the molecular complexity center. The bonds spanning the central atoms are used to make sets of strategic bonds that are the synthetic solutions proposed by the program strategy.

A reactivity predictor which evaluates the reactivity of the target (i.e., the group of routines which gives the real precursors, the type of reactions to be performed, the activating groups needed, etc.) is the part presently developed. One of its main objectives is the evaluation of strategic bond reactivity, which is mainly determined by their native polarity. As used here, NP is a quantitative characteristic of each bond in the target molecule from which the best functionalization of the precursors can be inferred.

A method for the determination of bond NPs in organic structures was already described in a recent paper.[3] The algorithm, based on the use of some molecular descriptors (atomic partial charges, atomic electronegativities, charge delocalization over conjugated bond systems, bond orders) and on the use of a logical, induced polarity for bonds with insufficient electronic definition (opposite polarity for the partner atom and for the vicinal atom; same polarity for the same atom involved in two successive bond breaks), assigns a polarity level to each considered bond, calculating the preferred charged state for each atom of the bond. Five charged states are possible for an atom: DEFINED (i.e., the activation level is good) positive or negative; SUGGESTED (i.e., more activation is needed, but the polarity is already decided) positive or negative; UNDEFINED (i.e., the activation level is low and the polarity is not definable). This procedure works correctly giving the expected results.

However, our continuous search for models useful in organic chemistry has given a different approach to the determination of bond NPs. This approach is described together with a new method for the calculation of the molecular electronic energy.

ELECTRONIC ENERGY

The description of a new procedure[5] for the calculation of molecular electronic energy is presented before the algorithm evaluating NPs. This procedure does not represent an alternative to the well-known quantomechanical methods. In our case, a much more approximated total energy is sufficient because the objective is a function discriminating states of the same molecule with quite different energies.

Our procedure originates from some ideas related to the work of Nalewajski[6] and Komorowski.[7] In recent years Nalewajski introduced the atom-in-molecule (AIM) approach.

His idea is to describe the molecular properties combining those of its component atoms. A similar approach has been elaborated by Komorowski. In a preceding paper,[8] we reported a method for the calculation of atomic and local hardnesses that is based on similar ideas; the method uses the well-known[9] relation between chemical potential and hardness, as described by

$$\eta = (\delta\mu/\delta N)_Z = -(\delta X/\delta N)_Z \tag{1}$$

where η is the hardness, μ is the chemical potential, N is the atom electronic population, X is the atom electronegativity, and Z is the atom core potential.

Using a modification of Gordy's[10] electronegativity[2b] (eq 2), it is possible to calculate the hardness as the derivative of

$$\mu = -X = -k_1 Z_{star}/R_{cov} + k_2 \tag{2}$$

the potential (eq 2') where R_{cov} is calculated using the corresponding Pauling's equation[11]

$$R_{cov} = R^0{}_{cov} Z^0{}_{star}{}'/Z_{star}{}'$$

obtaining

$$\mu = -X = -k_1 Z_{star} Z_{star}{}'/(Z^0{}_{star}{}' R^0{}_{cov}) + k_2$$

and finally

$$\eta = 0.35 k_1 (Z_{star} + Z_{star}{}')/(Z^0{}_{star}{}' R^0{}_{cov}) \tag{2'}$$

k_1 and k_2 are constants related to the position of the atom in the periodic table; Z_{star} is the Slater's effective core potential calculated using the Slater's relation

$$Z_{star} = Z - \sigma = Z - [N_1 + 0.85 N_2 + 0.35(N_3 - 1)]$$

where Z is the atomic nuclear charge, σ is the Slater's core screening factor, N_1 is the number of the inner-shell electrons, N_2 is the number of the medium-shell electrons, and N_3 is the number of the outer-shell electrons. $Z_{star}{}'$ is the Slater's effective core potential modified for a complete electron screening, i.e., where $(N_3 - 1)$ is replaced by N_3; R_{cov} is the covalent radius of the atom; the 0 superscript distinguishes the values of the isolated atom. N_3 varies depending on the environment of the atom in the molecule. A similar equation correlates the chemical potential to the electronic energy[6,7,9]

$$\mu = -X = (\delta E/\delta N)_Z \tag{3}$$

The calculation of the electronic energy is thus possible using eq 4, obtained by the integration of function 2 where

$$E = k_3(A + B + C) - k_2 N_3 \tag{4}$$

$$k_3 = -k_1/(Z^0{}_{star}{}' R^0{}_{cov})$$

$$A = (N^2 + aN - 2NN_1 - 2bNN_2 + N_1{}^2 + 2bN_1N_2 - aN_1 + b^2N_2{}^2 - abN_2)N_3$$

$$B = 0.5(-2aN + 2aN_1 + 2abN_2 - a^2)N_3{}^2$$

$$C = (a^{2/3})N_3{}^3$$

N is the total number of the electrons of the isolated atom; N_i is the occupation number of the ith shell. The constants a, b, c are the Slater's coefficients [respectively, a = 0.35 (0.3 for H atoms); b = 0.85; c = 1]. They are given as parameters to permit a successive optimization.[5] The other symbols have the same meaning as in the previous equations. The function obtained has a cubic dependence on electronic charge and is thus in agreement with the results of Komorowski.[7]

The simplification introduced in passing from function 3 to function 4 regards the invariability of the occupation of the inner shells. On physical grounds, this is acceptable as long as the molecular interaction is limited within the valence shell, as is the case for common interactions in organic chemistry.
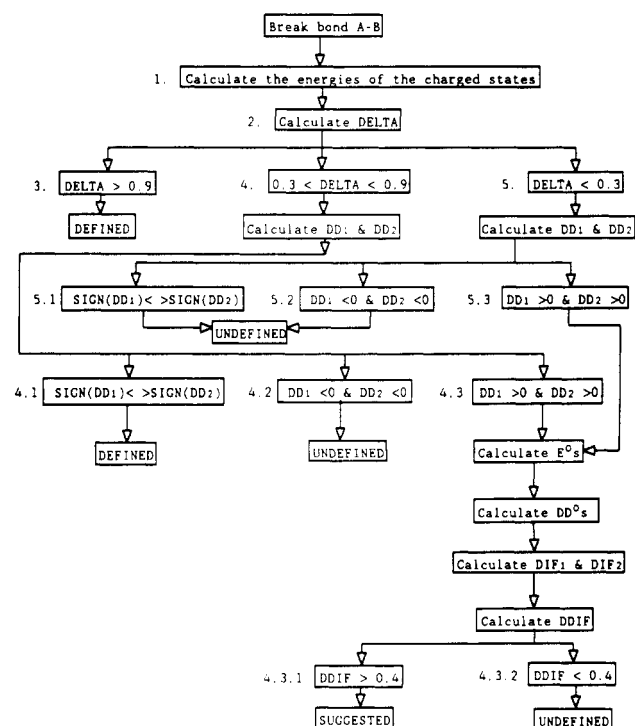
The method permits the calculation of the electronic energy of atoms. In agreement with a much more complex approach, mainly due to Bader,[12] the calculation of the molecular energy can be made summing the atomic contributions. It is obviously necessary to have available the atomic values corrected for the interatomic interactions. The dependence of the values of the shell occupation numbers on the atomic neighborhood in the molecule[2b-d] is reflected by the energy values, making possible the calculation for the atom in the molecule. Therefore, eq 5 is obtained to calculate the molecular electronic energy where

$$E = \sum_i E_i \tag{5}$$

the sum is over all the atoms in the molecule. Functions 4 and 5 are part of the following algorithm.

ALGORITHM

As the CAOS program Lilith considers only heteropolar reactions, the choice of the bond NPs is limited to the search for the best polarity between the two possible alternatives: minus charge on atom A and plus charge on atom B, or the reverse. The choice is made by selecting the best energy between the two calculated for the two charged structures (A^+/B^- or A^-/B^+).

The algorithm for the calculation of the bond NPs is based on a group of calculation procedures and a series of control steps which directs the flow of the program toward different blocks. The algorithm is also sketched in the flow chart of Scheme I, where there is reference to the following points.

The used thresholds (0.9[13] and 0.3 for DELTA, and 0.4 for DDIF as defined at points 2 and 4.3 below) permit the comparison of the present method with the previous one; they also help in the choice between different cases. However, they can also be completely ignored, and the procedure applied in its completeness to all cases. The results would be exactly the same.

Given a bond break between atom A and atom B:

(1) Four different molecular electronic energies are calculated for the charged molecules derived from the starting molecule with (a) A negative and B neutral [$E(A^-B^*)$], (b) A positive and B neutral [$E(A^+B^*)$], (c) A neutral and B

CALCULATION OF BOND NATIVE POLARITY

J. Chem. Inf. Comput. Sci., Vol. 32, No. 2, 1992 **127**

negative $[E(A^*B^-)]$, or (d) A neutral and B positive $[E-(A^*B^+)]$.

(2) The energy differences between 1a and 1d, and 1b and 1c are evaluated together with their difference DELTA $[\text{DELTA} = \text{ABS}\{[E(A^-B^*) - E(A^*B^+)] - [E(A^+B^*) - E-(A^*B^-)]\}]$.

(3) If DELTA is greater than the first threshold (0.9), the NP is set as DEFINED and the polarity (i.e., the combination of two of the preceding states) giving the greatest energy difference is chosen.

(4) If DELTA is smaller than 0.9 but greater than the second threshold (0.3), the analysis proceeds. The energy of the molecule with the bond broken but in the neutral (diradical) state is calculated $[E(A^*B^*)]$ and the differences between this last value and the energies of 1a (DD_1) and 1d (DD_2) are taken $[DD_1 = E(A^-B^*) - E(A^*B^*); DD_2 = E(A^*B^+) - E-(A^*B^*)]$. (If the largest DELTA is obtained for the difference between 1b and 1c, DD_1 and DD_2 would be calculated using 1b and 1c.) Then the signs of DD_1 and DD_2 are examined. In fact, DD_1 and DD_2 represent the (de)stabilization energies of the molecule going from the diradical to the charged state; the charged state is less stable than the neutral state if the sign is negative and more stable if the sign is positive.

Three cases are possible:

(4.1) The signs are discordant; one of the charged states is stable, and the polarity is DEFINED (e.g., again considering 1a and 1d the polarity is A^-/B^+).

(4.2) The signs are both negative. The charged molecules are less stable than the diradical, and thus the polarity cannot be set, i.e., it remains UNDEFINED.

(4.3) The signs are both positive. The charged molecules have similar stability. Another step is nevertheless available. It is possible to calculate another energy (E^0): the energy of the isolated atoms. This energy is calculated for the charged and the diradical molecules, and the differences between them and the corresponding ones, previously calculated, are taken.

$$DD^0_1 = E(A^-B^*) - E^0(A^-B^*)$$

$$DD^0_2 = E(A^*B^+) - E^0(A^*B^+)$$

$$DD^0_3 = E(A^*B^*) - E^0(A^*B^*)$$

These differences represent the (de)stabilization energies that are obtained going from separated atoms to molecules.

The differences between DD^0_1 and DD^0_3 (DIF_1) and DD^0_2 and DD^0_3 (DIF_2) are then calculated in order to compare with a common reference state. Finally their own difference DDIF $(\text{DDIF} = DIF_1 - DIF_2)$ is computed, and it is compared with a threshold (0.4). There are two possibilities:

(4.3.1) DDIF is greater than 0.4. One of the two charged molecules is sufficiently stable; therefore, it is possible to choose a polarity that will be SUGGESTED.

(4.3.2) DDIF is smaller than 0.4. None of the charged states is stable enough; the polarity remains UNDEFINED.

Returning to the DELTA calculated in step 2, the last possibility is:

(5) DELTA is smaller than the third threshold (0.3). This can happen if either the charged molecules $(A^-/B^+$ and $A^+/B^-)$ are both unstable giving an UNDEFINED polarity or one of the charged molecules with the same sign on both atom A and B $(A^-/B^-$ or $A^+/B^+)$ is the most stable.

To discriminate between the two cases, DD_1 and DD_2 are calculated as in step 4, but here the molecules compared are either 1a and 1c (both atoms negative) $[DD_1 = E(A^-B^*) - E(A^*B^*); DD_2 = E(A^*B^-) - E(A^*B^*)]$ or 1b and 1d (both atoms positive) $[DD_1 = E(A^+B^*) - E(A^*B^*); DD_2 = E-(A^*B^+) - E(A^*B^*)]$, depending on which of the two combinations has the best energy.

Three situations are possible:

Table I. Calculated Electronic Energies for Neutral and Charged Molecules[a]

compd	ab initio[b]	AM1[c]	engy	engy corr[d]
CH_4	39.98	6.73	13.59	
CH_3F	138.28	24.07	34.20	
CH_2F_2	236.61	41.41	54.74	
CHF_3	334.95	58.75	75.21	
CH_3CN	131.19	18.52	35.37	
$CH_2(CN)_2$	222.39	30.29	57.16	
$CH(CN)_3$	313.58	42.06	78.93	
CH_3NO_2	242.25	37.27	60.13	
$CH_2(NO_2)_2$	444.50	67.79	106.64	
$CH(NO_2)_3$	646.73	98.30	153.12	
CH_3^-	39.24	6.13	14.31	12.81
CH_2F^-	137.57	23.49	35.03	32.12
CHF_2^-	235.93	40.87	55.65	52.81
CF_3^-	334.32	58.26	76.19	73.35
CH_2CN^-	130.54	18.00	36.94	34.10
$CH(CN)_2^-$	221.82	29.83	58.92	52.28
$C(CN)_3^-$	313.07	41.64	80.78	71.71
$CH_2NO_2^-$	241.66	36.80	63.44	58.33
$CH(NO_2)_2^-$	443.97	67.39	112.86	102.85
$C(NO_2)_3^-$	646.23	97.93	162.30	147.13
CH_3^+	54.36	5.82	10.10	
CH_2F^+	101.78	23.18	30.60	
CHF_2^+	256.00	40.53	49.98	
CF_3^+	356.21	57.84	71.42	
CH_2CN^+	148.23	17.60	28.88	
$CH(CN)_2^+$	242.06	29.37	48.03	
$C(CN)_3^+$	335.68	41.13	67.25	
$CH_2NO_2^+$	260.32	36.31	54.68	
$CH(NO_2)_2^+$	653.08	66.76	100.00	
$C(NO_2)_3^+$	672.40	97.29	145.32	

[a] Energies are in atomic units, 1 au = 27.2116 eV. The label "engy" refers to our method. [b] Ref 14. Energy values for cationic molecules have been calculated using the HONDO8 program at the minimal basis set (STO3G). [c] Ref 15. [d] This column reports values for anionic molecules corrected taking into account the ratio between number of atoms and number of electrons.

(5.1) DD_1 and DD_2 have contrasting signs. A polarity can be chosen for only one atom, and at this level of the analysis this result is not sufficient to permit a choice. The polarity is UNDEFINED.

(5.2) Both DD_1 and DD_2 are negative. None of the atoms has a preferred charged state. The polarity is again UNDEFINED.

(5.3) Both DD_1 and DD_2 are positive. The same charge is stable on both atoms, and the possibility of favoring the umpolung of one or the other atom must be considered. The same analysis as described at point 4.3 is done, and similar considerations are possible. We can get two results: (1) DDIF is greater than 0.4; the umpolung of the charge of one atom is suggested (the polarity is also SUGGESTED) or (2) DDIF is smaller than 0.4; no atom umpolung can be suggested (the polarity is UNDEFINED, but in this case the two atoms will have the same charge).

The algorithm has been transformed into a computer routine, written in Fortran77 and implemented on a Bull X-20 (SVS Fortran compiler, Unix HBI version 2.40 O.S.).

RESULTS AND DISCUSSION

In order to verify the feasibility of the calculated energies, we have compared our values with both some published or calculated ab initio results,[14] and the energies calculated using the AM1[15] program. Tables I and II report the obtained values, and Table III reports the corresponding regression lines.

The first remark is on the different behavior of neutral and negatively charged structures; in fact, in our calculations anions are more stable than neutral molecules. The neglect of electron–electron interaction causes the problem. As a first consequence the regression lines are different for negatively charged and uncharged structures. On the other hand, the

128 *J. Chem. Inf. Comput. Sci., Vol. 32, No. 2, 1992*

BAUMER AND SELLO

Table II. Calculated Electronic Energies for Molecules Reported in Figure 1[a]

molecule	AM1[b]	engy	molecule	AM1[b]	engy
1	136.90	232.39	5	78.31	128.03
2	184.43	300.25	6	217.80	360.12
3	173.92	267.57	7	161.45	262.97
4	132.49	214.22	8	173.77	273.81

[a] Energies are in atomic units, 1 au = 27.2116 eV. The label "engy" refers to our method. [b] Ref 15.
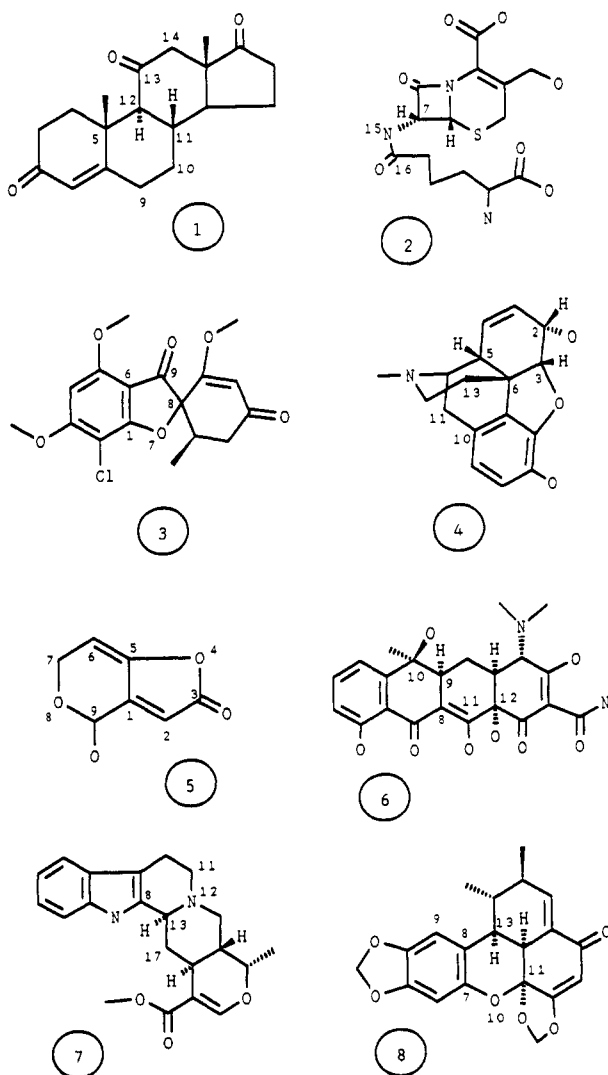
Figure 1. Structures for the molecules used in Table I reported with essential numbering.

introduction of a correction (that eliminates the differences related to the ratio between the number of electrons and the number of atoms in each molecule) permits the correlation of all the values on a single line (cf. Table III, lines 4 and 8).[16]

A second evidence comes from the correlation between AM1 and our results. Six different lines have been calculated for uncharged small molecules, negatively charged small molecules, positively charged small molecules, the complete set of small molecules with corrected energies, the set of the more complex molecules reported in Figure 1, and the set of all the uncharged molecules. It is therefore clear that a good correlation is always possible using the right coefficient.

A last remark is on the level of feasibility of each set of calculated energies. If we consider the ab initio values as the reference, we can affirm that our values correlate as well as or even better than AM1 values. Therefore, even remembering that our calculations are approximated, it is possible to use our results for the evaluation of the stabilities of molecules

Table III. Regression Lines and Coefficients of Energies Calculated by Ab Initio[a] (AB), AM1[b] (AM1), and Our Method (engy)

line no.	methods	line	coeff.
1	AB/engy uncharged	$Y = 3.98X$	0.99801
2	AB/engy charged(−)	$Y = 3.81X$	0.99673
		$(4.26)^c$	(0.99938)
3	AB[d]/engy charged (+)	$Y = 4.99X$	0.97738
4	AB/engy total	$Y = 4.47X$	0.97934
5	AM1/engy uncharged	$Y = 0.61X$	0.98069
6	AM1/engy charged(−)	$Y = 0.57X$	0.97911
		$(0.66)^c$	(0.98299)
7	AM1/engy charged(+)	$Y = 0.67X$	0.98955
8	AM1/engy total	$Y = 0.65X$	0.98725
9	AB/AM1 charged	$Y = 6.46X$	0.99087
10	AB/AM1 uncharged	$Y = 6.57X$	0.99120
11	AM1/engy Figure 1	$Y = 0.62X$	0.99643
12	AM1/engy uncharged + Figure 1	$Y = 0.61X$	0.99110

[a] Ref 14. [b] Ref 15. [c] The values in parentheses are obtained using corrected energies for (−)charged structures (see text). [d] Values for cations come from calculations made using HONDO8 program at the minimal basis set (STO3G).

Table IV. Bond Native Polarities Calculated Using Two Different Methods

structure	bond	NP[a]	NPat1[b]	NPat2[b]
1	13–14	1.215 (D)[c]	7.96 (D)[c]	−3.27 (D)[c]
	11–10	0.039 (U)	0.19 (U)	−0.19 (U)
	12–11	2.174 (D)	−9.15 (D)	−0.10 (U)
	12–5	0.541 (S)	−15.70 (D)	−14.60 (D)
	9–10	1.590 (D)	−7.29 (D)	−0.10 (U)
2	15–7	0.890 (S)	−16.21 (D)	−13.91 (D)
	15–16	0.377 (S)	−7.09 (D)	9.23 (D)
3	8–7	0.329 (S)	−20.90 (D)	−25.85 (D)
	9–6	0.532 (S)	6.81 (D)	−8.63 (D)
	7–1	2.726 (D)	−9.10 (D)	−2.22 (S)
4	3–2	0.463 (S)	−5.50 (D)	9.06 (D)
	10–11	0.648 (S)	−5.69 (D)	−0.31 (U)
	6–13	0.422 (S)	−0.35 (U)	−0.43 (U)
	6–5	0.193 (C)	−0.59 (U)	−0.22 (U)
5	6–7	0.609 (S)	−4.78 (S)	−3.18 (S)
	3–2	0.618 (S)	5.64 (D)	−7.93 (D)
	5–1	0.183 (C)	−5.43 (D)	−3.82 (S)
	8–9	1.726 (D)	−10.36 (D)	4.66 (S)
	4–5	3.289 (D)	−18.72 (D)	−7.98 (D)
6	12–11	1.256 (D)	−1.52 (S)	3.29 (S)
	10–9	0.709 (S)	−13.22 (D)	−7.48 (D)
	8–9	0.827 (S)	−9.88 (D)	−0.15 (U)
7	12–11	1.644 (D)	−6.90 (D)	1.20 (S)
	12–13	1.146 (D)	−4.31 (S)	−4.78 (S)
	13–17	0.586 (S)	−5.62 (D)	−0.60 (U)
	8–13	0.361 (S)	−3.26 (S)	−2.09 (S)
8	11–10	0.281 (S)	−13.43 (D)	−13.12 (D)
	8–13	0.638 (S)	−5.87 (D)	−0.13 (U)
	8–9	0.065 (C)	−2.66 (S)	−3.34 (S)
	7–8	0.466 (S)	5.98 (D)	−0.94 (U)

[a] Native polarities calculated by the method described in this paper. The order of the atoms in the bond column is such that the values in the NP column always refer to the first atom with (−) polarity. [b] Native polarities calculated by the method described in ref 3. Polarities refer to atoms, at1 being the first atom in the bond and at2 the second. [c] In parenthesis are reported the level of definition of the polarity: U = undefined, S = suggested, D = defined, and C = conflicting.

that have similar structures. This is exactly the case in the calculation of NPs.

The calculation of the NPs of the strategic bonds of some molecules have been made using the previous algorithm. (The corresponding molecular structures are drawn in Figure 1.) In Table IV the native polarities, calculated using both the present method and the one described in the preceding paper,[3] are reported. The examples have been chosen taking into particular account the needs of the discussion and do not intend to represent a complete set of results.

The comparison between the two methods is easily done by the direct inspection of Table IV; we will only comment on the chemical feasibility of the calculated polarities using standard chemical arguments.[17]

Molecule 1. In bond 13–14, 13 is negative because it is a carbonyl carbon. This contrasts with the usual experimental reactions which make bonds on carbonyl using the addition–substitution reaction. On the other hand, it is well-known that the aldehydic protons are acid,[18] and if a sufficiently strong base is used they give the corresponding anions followed by a rapid loss of carbon monoxide. Bond 11–10 has an UNDEFINED polarity because none of its atoms has near functional groups. In bond 12–11, 12 is negative because it has an α carbonyl. In bond 12–5, 12 is negative because it has an α carbonyl, whereas 5 has a γ carbonyl. In bond 9–10, 9 is negative because it has a γ carbonyl.

Molecule 2. In bond 15–7, 15 is negative because it is a nitrogen atom and it has an α carbonyl, whereas 7 has only an α carbonyl. In bond 15–16, 15 is again negative even if 16 is a carbonyl carbon (cf. atom 13 in molecule 1), because the charge is more stabilized on the electronegative heteroatom even if, in this break, it cannot be delocalized.

Molecule 3. In bond 8–7, 8 is negative because it has both an α and a γ carbonyl, therefore the stabilization given by the delocalization is more important than that given by the electronegativity of the oxygen atom 7. In bond 9–6, 9 is negative because it is a carbonyl carbon, whereas 6 is only an aryl carbon. In bond 7–1, 7 is negative because it is an oxygen atom, whereas 1 is only an aryl carbon.

Molecule 4. In bond 3–2, 3 is negative because it has an α oxygen atom, whereas 2 is also an allylic carbon and can stabilize a positive charge on the double bond. In bond 10–11, 10 is negative because it is an aryl carbon. In bond 6–13, 6 is negative because it is a benzyl carbon. In bond 6–5, 6 is again negative because 5 is only an allylic carbon.

Molecule 5. In bond 6–7, 6 is negative because it is a vinyl carbon and 7 has an α oxygen. In bond 3–2, 3 is negative because it is a carbonyl carbon, whereas 2 is only a vinyl carbon. In bond 5–1, both atoms are negative because both are vinyl carbons. In bond 8–9, 8 is negative because it is an oxygen, whereas 9 is only an allyl carbon. In bond 4–5, 4 is negative because it is an oxygen in a carboxyl group, whereas 5 is only a vinyl carbon.

Molecule 6. In bond 12–11, 12 is negative because it has an α carbonyl, whereas 11 is only a vinyl carbon. In bond 10–9, 10 is negative because it is a benzylic carbon, whereas 9 is only an allylic carbon. In bond 8–9, 8 is negative because it is a vinyl carbon and it has an α carbonyl.

Molecule 7. In bond 12–11, 12 is negative because it is a nitrogen. In bond 12–13, 12 is again negative because 13 is only a benzylic carbon. In bond 13–17, 13 is negative because it is a benzylic carbon. In bond 8–13, 8 is negative because it is an aryl carbon, whereas 13 has only an α nitrogen.

Molecule 8. In bond 11–10, 11 is negative because it has a γ carbonyl, whereas 10 is an oxygen; here again delocalization beats electronegativity. In bond 8–13, 8 is negative because it is an aryl carbon. In bond 8–9, both atoms are negative because they are both vinyl carbons. In bond 7–8, 7 is negative because it is a vinyl carbon and it has an α oxygen, whereas 8 is only a vinyl carbon.

Some common remarks follow.

First, if there is a contrast between delocalization on an electronegative atom and pure electronegativity, the negative charge goes on the atom in the first situation. Second, most of the possible ambiguous situations are solved by the method; beside the level of definition clearly mirrors the solved problem (e.g., a negative charge is more stable on a benzylic atom than on an allylic atom, but the energy difference is small and some

more activation is required). Third, the level of definition of the NPs diminishes passing from cases where only one of the atoms is functionalized to those where the electronic neighborhood is very similar.

It is also possible to sort the NPs obtaining a quantitative ordering. At the beginning of the list there are bonds with well-defined electronics (e.g., bond 4–5 in molecule 5, bond 7–1 in molecule 3, bond 12–11 in molecule 1); then there are solved ambiguities (e.g., bond 12–13 in molecule 7, bond 15–7 in molecule 2, bond 8–9 in molecule 6); at the end there are unsolved ambiguities (bond 6–5 in molecule 4, bond 5–1 in molecule 5) and bonds without activation (bond 11–10 in molecule 1).

Therefore, the results obtained with the present method permit the quantification of the concept of native polarity using a simple and straightforward algorithm based only on physical considerations. This could, in the future, furnish firm grounds for a possible expansion of the method to different aspects of molecular reactivity.

The use of bond NPs inside a CAOS program can help in the choice of the precursors of a target molecule, suggesting the functionalization of the target in dependence on the level of NP definition. Thus, for example, a well-defined negative atom can be obtained by a simple acid–base reaction (i.e., the precursor will only have one more H on that atom); whereas, a weaker negative atom requires more activation to be transformed in a nucleophile (i.e., the precursor will have an activating electron-withdrawing group on that atom). A solved ambiguity requires an umpolung on the atom that is less active; whereas, an unsolved ambiguity gives two possible umpolungs, one for each atom.

CONCLUSION

Two subjects have been presented in this paper: a new method for the calculation of the molecular electronic energy and a new procedure for the calculation of bond native polarities. The first method is an extension of the mathematical correlation of electronic energy, chemical potential, and hardness; it gives an effective means for evaluating a molecular quantity that can be easily used in comparing molecular states without resorting to heavy and cumbersome calculations. The second method establishes firm grounds for a previously realized fundamental part of our approach to synthesis design.

ACKNOWLEDGMENT

REFERENCES AND NOTES

(1) (a) Baumer, L.; Sala, G.; Sello, G. Organic Synthesis Planning: A New Algorithm for Strategic Bond Perception. *Tetrahedron* **1988**, *44*, 1195–1206. (b) Baumer, L.; Sala, G.; Sello, G. Organic Synthesis Planning: An Algorithm for Selecting Strategic Bond Forming Sequences. *Tetrahedron* **1989**, *45*, 2665–2676. (c) Baumer, L.; Sala, G.; Sello, G. Organic Synthesis Planning: Fundamental Principles in a New Approach (Lilith). *Anal. Chim. Acta* **1990**, *235*, 209–214.

(2) (a) Baumer, L.; Sala, G.; Sello, G. A Novel General Empirical Method for the Calculation of Atomic Components of Polarizability for Mean Molecular Polarizability Evaluation. *Gazz. Chim. Ital.* **1988**, *118*, 745–747. (b) Baumer, L.; Sala, G.; Sello, G. Residual Charges on Atoms in Organic Structures: A New Algorithm for Their Calculation. *Tetrahedron Comput. Methodol.* **1989**, *2*, 1–37. (c) Baumer, L.; Sala, G.; Sello, G. Residual Charges on Atoms in Organic Structures: A New Method for the Identification of Conjugated Systems and the Evaluation of Atomic Charge Distribution on Them. *Tetrahedron Comput. Methodol.* **1989**, *2*, 93–103. (d) Baumer, L.; Sala, G.; Sello, G. Residual Charges on Atoms in Organic Structures: Molecules Containing Charged and Backdonating Atoms. *Tetrahedron Comput. Methodol.* **1989**, *2*, 105–118.

(3) Baumer, L.; Sala, G.; Sello, G. Computer-Assisted Organic Synthesis Planning: Effective Bond Polarity as a Guideline to Reactivity. *J. Am. Chem. Soc.* **1990**, *113*, 2494–2500.

(4) Gasteiger, J.; Marsili, M.; Hutchings, M. G.; Saller, H.; Low, P.; Rose, P.; Rafeiner, K. Models for the Representation of Knowledge about Chemical Reactions. *J. Chem. Inf. Comput. Sci.* **1990**, *30*, 467–476.

(5) Asbrink, L.; Fridh, C.; Lindholm, E. HAM/3, a Semiempirical MO Theory. The SCF Method. *Chem. Phys. Lett.* **1977**, *52*, 63–68. This paper reports a similar approach for the calculation of electronic energies. This approach uses the semiempirical MO theory and represents a good precedent for our method.

(6) (a) Nalewajski, R. F. Recursive Combination Rules for Molecular Hardnesses and Electronegativities. *J. Phys. Chem.* **1989**, *93*, 2658–2666. (b) Nalewajski, R. F.; Korchowiec, J.; Zhou, Z. Molecular Hardness and Softness Parameters and Their Use in Chemistry. *Int. J. Quantum Chem., Quantum Chem. Symp.* **1988**, *22*, 349–366 and references cited therein.

(7) (a) Komorowski, L. Electronegativity and Hardness in the Chemical Approximation. *Chem. Phys.* **1987**, *114*, 55–71. (b) Komorowski, L. Chemical Hardness and L. Pauling Scale of Electronegativity. *Z. Naturforsch.* **1987**, *42a*, 767–773. (c) Komorowski, L. Empirical Evaluation of Chemical Hardness. *Chem. Phys. Lett.* **1987**, *134*, 536–540 and references cited therein.

(8) Baumer, L.; Sala, G.; Sello, G. A. New Method for the Calculation of Atomic and Local Hardness. *J. Comput. Chem.* **1990**, *11*, 694–699.

(9) (a) Pearson, R. G. Absolute Hardness: Companion Parameter to Absolute Electronegativity. *J. Am. Chem. Soc.* **1983**, *105*, 7512–7516. (b) Pearson, R. G. Absolute Electronegativity and Absolute Hardness of Lewis Acids and Bases. *J. Am. Chem. Soc.* **1985**, *107*, 6801–6806. (c) Yang, W.; Lee, C.; Ghosh, S. K. Molecular Softness as the Average of Atomic Softness: Companion Principle to the Geometric Mean Principle for Electronegativity Equalization. *J. Phys. Chem.* **1985**, *89*, 5412–5144. (d) Pearson, R. G. Absolute Electronegativity and Hardness Correlated with Molecular Orbital Theory. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 8440–8441. (e) Bohm, M. C.; Schmidt, P. C. Electronegativities and Hardnesses of the Mean Group Elements from

Density Functional Theory: Dependence on the Hybridization of the Chemical Bond. *Ber. Bunsen-Ges. Phys. Chem.* **1986**, *90*, 913–919. (f) Orsky, A. R.; Whitehead, M. A. Electronegativity in Density Functional Theory: Diatomic Bond Energies and Hardness Parameters. *Can. J. Chem.* **1987**, *65*, 1970. (g) Pearson, R. G. Chemical Hardness and Bond Dissociation Energies. *J. Am. Chem. Soc.* **1988**, *110*, 7684–7690. (h) Pearson, R. G. Absolute Electronegativity and Hardness: Applications to Organic Chemistry. *J. Org. Chem.* **1989**, *54*, 1423–1430.

(10) (a) Gordy, W. A. New Method of Determining Electronegativity from Other Atomic Properties. *Phys. Rev.* **1946**, *69*, 604–607. (b) Gordy, W.; Thomas, W. J. O. Electronegativities of the Elements. *J. Chem. Phys.* **1956**, *24*, 439–444.

(11) Pauling, L. *The Nature of the Chemical Bond*, 3rd ed.; Cornell University: Ithaca, NY, 1960.

(12) Bader, R. F. W. Atoms in Molecules. *Acc. Chem. Res.* **1985**, *18*, 9 and references cited therein.

(13) Thresholds are energy differences and could have a dimension, but they are used dimensionless to free them from any precise physical meaning; therefore maintaining a relation with the thresholds used in the previous approach.[3]

(14) (a) Edgecombe, K. E.; Boyd, R. J. Molecular Orbital Treatment of Substituent Effects. I. Structures of Some Carbon Acids and Their Conjugated Bases. *Can. J. Chem.* **1983**, *61*, 45. (b) Ab initio results for cations were obtained using the program HONDO8 (included in the MOTECC package) at the minimal basis set (STO3G).

(15) Ampac: Austin Method 1 Package; QCPE Program No. 527.

(16) As our algorithm calculating NPs works only with energy differences, this correction is not required.

(17) The chemical evidence is mostly taken from: March, J. *Advanced Organic Chemistry*, 3rd ed.; Wiley-Interscience: New York, 1985.

(18) Chandrasekhar, J.; Andrade, J. G.; Schleyer, P. von Ragué Efficient and Accurate Calculation of Anion Proton Affinities. *J. Am. Chem. Soc.* **1981**, *103*, 5609–5612.

Generation of Molecular Graphs for QSAR Studies: An Approach Based on Acyclic Fragment Combinations

S. S. TRATCH, O. A. LOMOVA, D. V. SUKHACHEV, V. A. PALYULIN, and N. S. ZEFIROV*

Department of Chemistry, Moscow State University, Moscow 119899, USSR

A generating algorithm for substituted derivatives of a given structure is elaborated for the purpose of QSAR studies, and the complex substituents are, in turn, constructed from given sets of elementary fragments. Elementary as well as composite fragments are classified into terminal, linear, and branched ones; three types of combining operations are introduced for these fragments. The combining operations make it possible to correctly generate the complete sets of substituents with prescribed values of several numerical characteristics. A detailed description of the generating procedure is presented. The application of graph and permutation group theories shows that only automorphism groups of some elementary branched fragments are required for nonduplicate construction of all substituents. The related analytical enumeration problems are solved on the basis of Burnside's Lemma. Additional selection criteria and internal representations of fragments are also discussed. The WLN codes of fragments are actually used in computer implementation of the suggested algorithm.

1. INTRODUCTION

During the last decade the interest in the problem of computer-aided design of organic structures with prescribed properties has significantly increased. The investigation of quantitative structure–activity relationships (QSAR) has become a popular and often powerful method for predicting many kinds of biological activity.[1,2] In the most typical situation, the specific parent structure together with a certain set of substituents is initially found to be responsible for the desired activity. At the second stage, some structure generator is needed to search for new active derivatives of the parent "skeleton". It is significant that the enlarged set consisting of generated target structures can lead to the refinement of the initial structure–activity model owing to detection of new highly favored or disfavored substituents.

The well-known generation techniques are oriented, however, either to the constructive enumeration of nonisomorphic abstract graphs[3–5] or to the construction of acyclic[6–8] or cyclic[9–16]

molecular graphs corresponding to a given gross formula. Such generators seem to be quite inadequate for the QSAR studies because not the gross formula but the parent skeleton must be present in all generated target structures. On the other hand, the broad class of relatively complex substituents (having been, in turn, constructed of atoms and simple groups) seems to be preferable as compared to the fixed set of given substituents. It results in some similarity between desired and "normal" graph generators; in particular, all generators must produce only nonduplicate target structures with the total number of structures being dependent on certain limitations.

In this paper, we describe our efforts to elaborate the fragment generator which makes it possible to sequentially construct all *acyclic* substituents from a given set of "elementary" fragments. The idea of the suggested algorithm is somewhat similar to that used in ref 11, but in our approach the elementary fragments are allowed to be represented not only by atoms but also by arbitrary multiatomic groups. The