electron transfer process, but are the result of spontaneous interaction between acetylene molecules and alkali metal atoms in their excited ^2P states. Formation of a certain amount of the Li-acetylene complex in the dark is attributed to a much higher temperature required for vaporization of Li atoms, and much greater mobility of the latter in a fluid surface layer of the matrix during deposition.

The ESR spectrum of well-annealed Li-acetylene complex was thus observed when the matrix was simultaneously irradiated with orange light (λ = 650~700 nm) during deposition, and the K-acetylene complex was formed only after irradiation of the matrix with red light (λ = 750 ± 50 nm). In both cases the isomerization to the vinylidene form was observed to occur on exposure to light of $\lambda \leq 600$ nm. The situation in the Na/acetylene/argon system must be that the π complexes are formed upon irradiation of the matrix with yellow light (λ = 600 ± 50 nm), but are immediately isomerized to the vinylidene form by the ensuing radiation.

As stated earlier, the alkali metal atom-acetylene complex is essentially a charge-transfer complex, a bent acetylene anion interacting side-on with an alkali metal cation. The photoenergy of $\lambda \leq 600$ nm (= 48 kcal/mol) required for its isomerization to the vinylidene form is in excellent agreement with the theoretically predicted barrier height of 40~50 kcal/mol for the process.[3,5,6] The electronic transition involved here must be that which correlates to the $\pi_z \rightarrow \pi_y^*$ transition of the acetylene moiety. The $\pi \rightarrow \pi^*$ transition of an isolated acetylene occurs at ~240 nm.[23] A rather large red-shift of the transition to ~600 nm in the π complex is not unreasonable in view of the distortion and the interaction involving the π_y^* orbital.

Acknowledgment. My thanks to S. Sakai and M. T. Nguyen for stimulating discussions of the subject.

(23) Herzberg, G. *Electronic Spectra of Polyatomic Molecules*; D. Van Nostrand: New York, 1967; p 611.

A New Definition of Functional Groups and a General Procedure for Their Identification in Organic Structures

Guido Sello

Contribution from the Dipartimento di Chimica Organica e Industriale, Universita' degli Studi di Milano, via Venezian 21, I-20133 Milano, Italy. *Received June 3, 1991*

Abstract: Functional groups (FGs) have always been used to describe structures and reactivities of organic molecules, therefore their univocal identification is important. Current approaches are often subject to limits and errors, using descriptions that are either too approximated or too rigid. A new definition, and the corresponding search procedure, based on the calculation of a molecular descriptor is introduced and applied to some molecules of different complexity. Results show a consistent set of FGs for each structure. They are divided into two classes—first level FGs and second level FGs—that describe interactions either among atoms (first level) or among groups of atoms (second level).

Introduction

The recognition of functional groups (FGs) in organic molecules is important both for their handling (e.g. in organic reactivity modeling) and for their storage and retrieval in reaction databases. FG identification is also important in all computer programs for organic synthesis planning, both those using a database of reactions and those using a mechanicistic approach (in this last case the recognition can be implicit).

The inspection of an organic structural formula by a chemist involves the direct and efficient perception of the molecular characteristics through a mental mechanism that is a complex collaborative interaction between symbolic and graphical recognition and memory correlation. A similar operation carried out by a computer requires great care in the exact definition of the object to be perceived and in the realization of a general procedure that can be applied to any molecule.

There are two possible approaches to the problem: (1) the choice of a set of fundamental FGs that permits the recognition of the FGs in a given molecule via an accurate comparison between its atomic groups and the FG set;[1,2] (2) the definition of a set of rules, listing the necessary requisites of an FG, that can be applied to a given molecule furnishing its FGs.

The first approach is the most similar to the chemist's style: his/her knowledge is represented by the chosen FG set and his/her mental process is simulated by the comparison procedure. Both the chemist and the machine will take advantage of the exact definition of the FGs, thus the analysis will be fast and reliable. On the other hand, both the chemist and the machine will suffer the limited number of FGs in the set and, for some particular structures, the FG recognition could be partial or erroneous. All the current methods for FG identification are of this type.

The second approach will be generally applicable (if the rules are sufficiently comprehensive), permitting the identification of uncommon and even new FGs; but it will suffer a longer search time and, mainly, the possible insertion of apparently similar atomic groups into different FG classes.

The procedure being presented is of the second approach type that guarantees a greater potential applicability. In addition the procedure is based on the calculation of a molecular descriptor that is a number and not a descriptive representation; therefore it can be an effective aid in determining a correct FG classification, useful in describing and handling molecules. Eventually it is in accord with our approach to synthesis design[3] that does not use a priori defined FGs. The main results are as follows: (1) the possibility of identifying FGs in any molecule, regardless of its complexity; (2) the independence of the FGs from an a priori

(1) Esack, A.; Bersohn, M. A Program for Rapid and Automatic Functional Group Recognition. *J. Chem. Soc., Perkin Trans. I* **1974**, 2463-2470. For an original approach to the representation of organic reactions and reaction centers see also: Fujita, Shinsaku Description of Organic Reactions Based on Imaginary Transition Structures. 9. Single-Access Perception of Rearrangement Reactions. *J. Chem. Info. Comput. Sci.* **1987**, *27*, 115-120 and references cited therein.

(2) Corey, E. J.; Wipke, W. T.; Cramer, R. D.; Howe, W. J. Techniques for Perception by a Computer of Synthetically Significant Structural Features in Complex Molecules. *J. Am. Chem. Soc.* **1972**, *94*, 431-439.

(3) Baumer, L.; Sala, G.; Sello, G. Organic Synthesis Planning: Fundamental Principles in a New Approach (Lilith). *Anal. Chim. Acta* **1990**, *235*, 209-214.

0002-7863/92/1514-3306$03.00/0

specified appearance, therefore permitting the automatic individualization of new FGs without changing the procedure; (3) the quantification of the FG definition that permits its use in handling FG reactivity; and (4) the possibility of calculating other FG properties (e.g. local hardness) using component atoms that are really interacting.

Molecular Electronic Energy

In a previous paper[4] we describe a method for the calculation of the molecular electronic energy. The method uses the well-known relation[5] between electronic energy and chemical potential, as reported in eq 1,

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

where μ is the chemical potential, X is the atomic electronegativity, E is the electronic energy, N is the atom electronic population, and Z is the atom core potential that is considered constant in the derivative.

By using a modification of Gordy's electronegativity,[6] it is possible to calculate the hardness[7] by derivation and the electronic energy by integration.[8] The method permits the calculation of the electronic energy for each atom. In agreement with a much more complex approach, mainly due to Bader,[9] the calculation of the molecular energy can be made by summing the atomic contributions (eq 5)

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

where the sum is over all the atoms in the molecule and E_i is the energy of atom i.

The Search for FGs

A general and exact definition of a functional group is absent in the chemical literature, whereas a descriptive or approximate definition is much more common.[10] Otherwise, the possibility of calculating a descriptor that certifies an atom as an FG component permits an exact definition of an FG. The descriptor must be sensitive to the structure of the molecule to which the atom belongs; such a descriptor can be the electronic energy.

The electronic energy of a molecule depends on the component atoms and their interactions. In fact, if an atom has good interactions it stabilizes the molecule. In the molecular orbital sense, the combination of the atomic orbitals gives stable molecular orbitals if the attractive energies (core–electron) of the atoms are greater than the repulsive energies (both core–core and electron–electron), i.e. if the interaction of the electrons of one atom with the other nucleii stabilizes the molecule more than the interactions between the nucleii and/or between the electrons destabilize it.

If the energetic contribution of an atom pair is high, i.e. if it highly stabilizes the molecule, we can say that the two atoms have a "strong" interaction. Atoms that have strong interactions are qualified to be FG components, i.e. they are "important". The "importance" of an atom is, in this context, a quantity; it is thus possible to find the most "important" atoms and to call them "central" atoms.

An FG is thus composed of a set of connected atoms that are considered sufficiently "important"; we only need the rule to distribute the atoms in different FGs. If we consider three connected atoms A-B-C, which respectively have E_A, E_B, and E_C energy contributions, we can have four different situations:

(1) $E_A > E_B > E_C$—the importance is decreasing; all the atoms belong to the same FG.

(2) $E_A > E_B < E_C$; $E_A > E_C$—the importance is initially decreasing and then increasing; atom A is more important than atom C, A and B belong to one FG, and C belongs to another FG.

(3) $E_A > E_B < E_C$; $E_A < E_C$—the importance is initially decreasing and then increasing; atom C is more important than atom A, A belongs to one FG, and B and C belong to another FG.

(4) E_B is too low, therefore B is not an FG atom—A and C belong to two different FGs.

It is thus possible to give the following definition: "an FG is a set of sufficiently important connected atoms, in which the importance is always decreasing from the central atom towards the peripheral atoms", where "importance" is equivalent to "contribution to the electronic stabilization energy of the molecule".

The just defined FGs are called "first level functional groups" (FLFGs). It is also possible to define "second level functional groups" (SLFGs) obtained by the combination of FLFGs that have one or more FG atoms in common.

Experimental Procedure

The procedure that actually searches and finds FGs is a straightforward application of the principles outlined in the previous section.

Its main activity is the calculation of the molecular energies for the given molecule (T) and for all the molecules obtained from T by isolating in turn one non-hydrogen atom from the others; the isolation of each atom is accomplished by cutting all the bonds connecting it to its neighbors, thus eliminating all the corresponding interactions.

Having the energy set available, the search for the FGs can start.

(1) Difference Calculation. The comparison of the relative "importance" of each atom requires a set of homogeneous values. This set is composed of the (de)stabilization energies ($DESEN_i$) of all the molecules obtained from T. The energies are calculated by taking the difference between the electronic energy of each molecule and that of the reference T ($DESEN_i = DELTA_i - DELTA_T$). All the energies (T energy included) are weighted against the sum of the energies of all the isolated atoms ($DELTA_i = ENERGY_i - ENERGY_0$).

(2) FLFG Search. The first level functional groups are obtained by collecting together "important" atoms that can be isolated from the molecule (the "importance" is obtained by comparison with a threshold THR1 equal to 0.02). The search begins with the location of the most "important" FG atoms; then the atoms directly connected to it are examined and, if they qualify as FG atoms, inserted in the FG.

The search continues until either no more FG atoms are found or the importance of all the examined atoms is increasing.

In this last case a check is also made to verify if the FG atoms found in the previous step really belong to the present FG. The check consists of the comparison between the DESENs of the last examined atoms and those of the atoms examined two steps before. If the DESEN difference is greater than a second threshold (THR2 equal to 1.0), then the atoms found in the previous step are eliminated from the present FG.

The use of two different thresholds is due to the assumption that an FG atom can belong to two different FLFGs only if its neighbors are not very important.

Two other types of FGs are identified to complete the FG set. In fact, there can be two classes of atoms that are not energetically "important", but that represent a potential source of "perturbation" of the molecule. The first class is formed by the atoms with a great difference in their hardnesses (this characteristic is a well-known source of reactivity). The second class is formed by atoms connected to reactive hydrogens; most of them are also energetically "important", but it may happen that in highly delocalized molecules the (de)stabilization energy is distributed over the entire conjugated system and thus its magnitude is small for each atom.

The procedure is repeated on all the central atoms. At the end, all the FLFGs are identified.

(3) SLFG Search. Having the FLFGs available, it is possible to build the second level functional groups. They are FGs composed of pairs of

(4) Baumer, L.; Sello, G. A New Method for the Calculation of Bond Natural Polarity Using Molecular Electronic Energy. *J. Chem. Info. Comput. Sci.* **1991**, submitted.

(5) Parr, R. G.; Pearson, R. G. Absolute Hardness: Companion Parameter to Absolute Electronegativity. *J. Am. Chem. Soc.* **1983**, *105*, 7512–7516.

(6) (a) Baumer, L.; Sala, G.; Sello, G. Residual Charges on Atoms in Organic Structures: A New Algorithm for Their Calculation. *Tetrahedron Comput. Methodology* **1989**, *2*, 1–37. (b) 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. Methodology* **1989**, *2*, 93–103. (c) Baumer, L.; Sala, G.; Sello, G. Residual Charges on Atoms in Organic Structures: Molecules Containing Charged and Backdonating Atoms. *Tetrahedron Comput. Methodology* **1989**, *2*, 105–118.

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

(8) A more complete description of the calculation of molecular electronic energy is available as supplementary material.

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

(10) Examples of FG definitions are present in the following: (a) Gold, V. Glossary of Terms Used in Physical Organic Chemistry. *Pure Appl. Chem.* **1983**, 1281–1371. (b) McMurry, J. *Fundamentals of Organic Chemistry*; Wadsworth, Inc.: Belmont, CA, 1986. (c) Reusch, W. H. *An Introduction to Organic Chemistry*; Holden-Day, Inc.: San Francisco, CA, 1977.

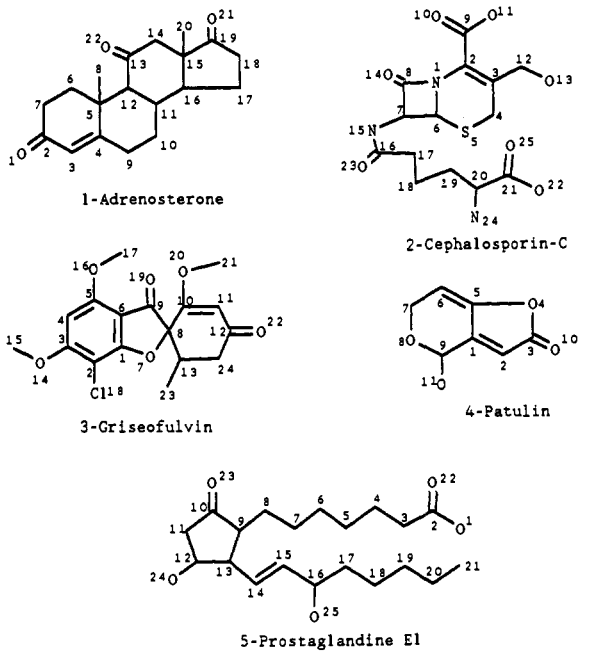

Figure 1. Atom numbering for structures **1–5**.

Table I. (De)stabilization Energies of Non-H Atoms in Adrenosterone[a]

atom no.	energy[b]	atom no.	energy[b]	atom no.	energy[b]
1	5.637	8	0.017	16	0.006
2	5.318	9	0.283	17	0.012
3	1.560	10	0.012	18	0.245
4	1.814	11	0.006	19	5.538
5	0.297	12	0.253	20	0.017
6	0.012	13	5.540	21	5.538
7	0.247	14	0.246	22	5.540
		15	0.261		

[a] Atom numbers refer to Figure 1. [b] Energy is in a.u.

FLFGs with one or more common atoms.

A last set of SLFGs is formed by the combination of all the SLFGs that have at least one common atom. In this case the SLFGs are made up of two or more SLFGs and are the maximum overlap of the FGs.

Results and Discussion

An application of the procedure will follow in order to make its functioning more clear. The chosen molecule is adrenosterone (Figure 1), which is simple enough to be accurately described.

The first step is the calculation of the (de)stabilization energies for all the non-H atoms. These are reported in Table I. Atoms 6, 8, 10, 11, 16, 17, and 20 are not "important" enough and they are not part of any FG. It must be noted that the energies for atoms with a similar molecular neighborhood are comparable (e.g. note carbonyl carbons, unsaturated carbons, and α-carbonyl carbons) and can give some hints on different reactional behavior even for quite similar groups (e.g. carbon 2 (unsaturated carbonyl) compared with carbons 13 and 19 (saturated carbonyls)).

Then the procedure begins to build FGs.

(1) Starting from atom 1, an FG composed of atoms 1, 2, 3, and 7 is determined. The search is interrupted on atom 6 (non-FG atom) and on atom 4 (an atom with an energy greater than atom 3).

(2) Starting from atom 4, an FG composed of atoms 3, 4, 5, 9, and 12 is determined. The search is interrupted on atoms 6, 8, and 10 (non-FG atoms) and on atoms 2 and 13 (atoms with energies greater than atoms 3 and 12).

(3) Starting from atom 22, an FG composed of atoms 12, 13, 14, and 22 is determined. The search is interrupted on atom 11 (non-FG atom) and on atoms 5 and 15 (atoms with energies greater than atoms 12 and 14).

(4) Starting from atom 19, an FG composed of atoms 14, 15, 18, 19, and 21 is determined. The search is interrupted on atoms 16, 17, and 20 (non-FG atoms) and on atom 13 (an atom with an energy greater than atom 14).

(5) Atoms 2, 3, 5, 7, 9, 12, 13, 14, 15, 18, and 21 are not used as starting atoms because they are not central FG atoms.

At this point the search is finished and the FGs are examined to ensure that their atoms do not belong to another FG. Atom 12 in FG no. 2 and atom 14 in FG no. 4 are therefore eliminated from these two FGs and remain only in FG no. 3.

Finally the search for SLFGs is undertaken. Only FG nos. 1 and 2 can be combined, giving FG no. 5 composed of atoms 1, 2, 3, 4, 5, 7, and 9.

The discussion is now extended to the set of 10 molecules reported in Figures 1 and 2. Their FGs are given in Table II and III, and the FG atom energies are given in Tables IV and V. The results will be discussed in turn below.

Table II. Functional Groups Identified in the Molecules of the Set[a]

molecule	FLFG	SLFG
1	1-2-3-7	1-2-3-4-5-7-9
	3-4-5-9	
	12-13-14-22	
	15-18-19-21	
2	1-2-3-4-6-12	1-2-3-4-5-6-12
	4-5-6	1-2-3-4-6-7-8-12-14
	1-6-7-8-14	1-2-3-4-6-12-13
	12-13	1-2-3-4-6-9-10-11-12
	1-2-6-9-10-11	1-4-5-6-7-8-14
	15-16-17-23	1-2-4-5-6-9-10-11
	20-21-22-25	1-2-6-7-8-9-10-11-14
	20-24	20-21-22-24-25
		1-2-3-4-5-6-7-8-9-10-11-12-13-14
3	1-2-7-18	1-2-3-4-7-14-18
	2-3-4-14-18	1-2-7-8-10-11-18-20
	4-5-16	1-2-6-7-8-9-18-19
	7-8-10-11-20	2-3-4-5-14-16-18
	16-17	2-3-4-14-15-18
	14-15	4-5-16-17
	6-7-8-9-19	6-7-8-9-10-11-19-20
	20-21	7-8-10-11-20-21
	11-12-22-24	7-8-10-11-12-20-22-24
		1-2-3-4-5-6-7-8-9-10-11-12-14-15-16-17-18-19-20-21-22-24
4	7-8	1-4-5-6-7-8-9
	1-4-5-6-7-9	1-2-3-4-5-6-7-9-10
	1-2-3-4-9-10	1-4-5-6-7-9-11
	9-11	1-2-3-4-9-10-11
		1-2-3-4-5-6-7-8-9-10-11
5	1-2-3-22	9-10-11-12-13-14-15-16-23
	9-10-11-12-23	9-10-11-12-23-24
	12-13-14-15-16	12-13-14-15-16-24
	12-24	12-13-14-15-16-25
	16-25	9-10-11-12-13-14-15-16-23-24-25
6	1-2-3-4-5-6-7-8-9-10-11-12-23	1-2-3-4-5-6-7-8-9-10-11-12-13-14-18-23
		1-2-3-4-5-6-7-8-9-10-11-12-22-23
	12-13-14-18	1-2-3-4-5-6-7-8-9-10-11-12-23-24
	14-15-20-21	12-13-14-15-18-20-21
	11-12-22	11-12-13-14-18-22
	23-24	1-2-3-4-5-6-7-8-9-10-11-12-13-14-15-18-20-21-22-23-24

[a] Atom numbers refer to Figures 1 (structures **1–5**) and 2 (structure **6**).

Adrenosterone (1). This molecule has been analyzed already; its FGs are simple and correspond to the usual chemist's perception. It is noticeable that the conjugated carbonyl is found only at the second level while the system is divided into a carbonyl and a double bond at the first level. Atom 5 is part of the unsaturated carbonyl, even if it has no active H's; in fact, it is potentially reactive (e.g. in a possible rearrangement) because of its allylic position.

Cephalosporin-C (2). There are 8 FLFGs in the molecule corresponding to the main FGs. FG no. 2 is found because its atoms have different hardnesses. The allylic alcohol group (2-3-12-13) is not found as an FLFG, while the enamine group is

Table III. Functional Groups Identified in the Molecules of the Set[a]

molecule	FLFG	SLFG
7	3-4-5-6-10	3-4-5-6-8-9-10-11-12-23
	8-9-10-11-12-23	2-3-4-5-6-10-19
	12-15-16-17-18-25	3-4-5-6-10-21
	2-3-19	1-3-4-5-6-7-8-9-10-22
	10-21	8-9-10-11-12-15-16-17-18-23-25
	1-7-8-9-10-22	8-9-10-11-12-21-23
	12-24	1-7-8-9-10-11-12-22-23
	16-17-18-26-27-28	8-9-10-11-12-23-24
	17-18-29	12-15-16-17-18-24-25
	18-30-31-32	12-15-16-17-18-25-26-27-28
		12-15-16-17-18-25-29
		12-15-16-17-18-25-30-31-32
		1-7-8-9-10-21-22
		16-17-18-26-27-28-29
		16-17-18-26-27-28-30-31-32
		17-18-29-30-31-32
		1-2-3-4-5-6-7-8-9-10-11-12-15-16-17-18-19-21-22-23-24-25-26-27-28-29-30-31-32
8	3-4-5-6-16	3-4-5-6-16-17
	9-10-11-12-14	9-10-11-12-13-14
	12-13	
	16-17	
9	2-3-4	1-2-3-4-5-6-8-9-13
	1-2-4-5-6-8-9-13	1-2-4-5-6-7-8-9-10-13
	6-7-8-10-13	1-2-4-5-6-8-9-12-13-15-16-17
	12-13-15-16-17	1-2-4-5-6-8-9-12-13-17-18-23-26
	11-19-20-26	6-7-8-10-12-13-15-16-17
	21-22	6-7-8-10-12-13-17-18-23-26
	12-13-17-18-23-26	12-13-15-16-17-18-23-26
		11-12-13-17-18-19-20-23-26
		1-2-3-4-5-6-7-8-9-10-11-12-13-15-16-17-18-19-20-23-26
10	1-2	1-2-3-4-5-6-9-35
	10-11-12-13-14-15-17-37-38	10-11-12-13-14-15-17-18-19-37-38
		10-11-12-13-14-15-17-37-38-39
	15-18-19	16-20-21-22-23-40-42
	16-20-21-40	16-20-21-40-42
	20-22-23-42	20-22-23-24-25-42-43
	26-27-28-29	26-27-28-29-30
	29-30	26-27-28-29-44
	31-32-33-34	2-3-4-5-6-7-8-9-35-36
	2-3-4-5-6-9-35	1-2-3-4-5-6-7-8-9-35-36
	7-8-9-36	10-11-12-13-14-15-17-18-19-37-38-39
	38-39	16-20-21-22-23-24-25-40-41-42-43
	40-41	26-27-28-29-30-44
	23-24-25-43	
	26-44	

[a] Atom numbers refer to Figure 2.

Table IV. (De)stabilization Energies of FG Atoms[a]

molecule	atom	energy[b]	atom	energy[b]	atom	energy[b]
1	1	5.637	2	5.318	3	1.560
	4	1.814	5	0.297	7	0.247
	9	0.283	12	0.253	13	5.540
	14	0.246	15	0.261	18	0.245
	19	5.538	21	5.538	22	5.540
2	1	0.514	2	1.596	3	1.828
	4	0.272	5	0.312	6	0.027
	7	0.235	8	5.589	9	5.325
	10	5.548	11	0.497	12	0.179
	13	0.790	14	5.574	15	0.379
	16	5.544	17	0.237	20	0.231
	21	5.542	22	0.499	23	5.533
	24	0.471	25	5.470		
3	1	0.510	2	0.363	3	0.394
	4	0.343	5	0.415	6	0.159
	7	0.089	8	0.446	9	5.454
	10	1.822	11	1.472	12	5.313
	14	0.059	15	0.150	16	0.060
	17	0.152	18	0.239	19	5.695
	20	0.123	21	0.152	22	5.634
4	1	1.495	2	1.551	3	5.380
	4	0.520	5	1.525	6	1.705
	7	0.184	8	0.234	9	0.109
	10	5.588	11	0.778		
5	1	0.515	2	5.546	3	0.232
	9	0.253	10	5.555	11	0.246
	12	0.102	13	0.313	14	1.747
	15	1.747	16	0.218	22	5.455
	23	5.555	24	0.806	25	0.805
6	1	0.912	2	0.742	3	0.770
	4	0.778	5	0.844	6	1.673
	7	0.616	8	0.558	9	1.644
	10	1.714	11	0.136	12	0.021
	13	0.058	14	0.028	15	0.313
	18	0.028	20	1.766	21	1.772
	22	0.795	23	0.050	24	0.163
7	1	0.254	2	0.440	3	0.357
	4	0.476	5	0.499	6	0.496
	7	5.261	8	1.437	9	0.301
	10	0.144	11	1.792	12	0.470
	15	5.549	16	0.501	17	0.100
	18	0.021	19	0.489	21	0.782
	22	5.813	23	0.500	24	0.774
	25	5.549	26	5.540	27	5.532
	28	0.439	29	0.789	30	0.061
	31	0.034	32	0.034		

[a] Atom numbers refer to Figures 1 (structures **1**–**5**) and 2 (structures **6** and **7**). [b] Energy is in a.u.

part of the complex group (1-2-3-4-6-12) of the double bond. The β-lactam ring (1-6-7-8-14) is an FLFG as the aminoacidic part (1-2-6-9-10-11) attached to the six-membered ring, while the aminoacidic part of the chain (20-21-22-24-25) is identified as an acid plus an amine.

The SLFGs are more and more complex as the combination procedure proceeds until the most complex FG (1-2-3-4-5-6-7-8-9-10-11-12-13-14), which comprises the two fused rings, is found. Among SLFGs there are the following: (a) 3 combinations that include the β-lactam ring, showing its connections with the double bond 2-3, the sulfide group 4-5-6, and the acid 9-10-11—it is otherwise interesting to note that the lactam is not connected with the chain amide 15-16-23; (b) the complete six-membered ring; (c) the enamine combined with the allylic alcohol and with acid 9-10-11; (d) the acid 9-10-11 combined with the sulfide; and (e) the acid 21-22-25 combined with the amine 20-24. Atoms 17-18-19 (non-FG atoms) separate the tail aminoacid from the rest of the molecule, and bond 7-15 (7 interacting mainly with 8 and 15 with 16) separates the chain amide from the ring system.

Griseofulvin (3). There are 9 FLFGs in the molecule. Here it is interesting to observe the subdivision of the aromatic ring system. It is divided into 4 FLFGs (1-2-7-18, 2-3-4-14-18, 4-5-16, 6-7-8-9-19), showing some atom interactions (7-18, 14-18, 7-9 through 8 and not through 1).

SLFGs contain most of the possible connections among the groups present in the molecule: (a) the aromatic ring is still divided into pieces that now show longer range interactions (meta–para positions on the ring, or even through the spiro atom to the enol group 10-11-20); (b) the oxodihydrofuran is a self-standing FG; (c) enols and ketones combine in every possible way and include atoms 7-8 in each combination; and (d) atoms 7 and 8 are present in 6 SLFGs from a total of 9, the 3 excluded being subdivisions of the left half of the aromatic ring.

The entire molecule (excluded atoms 13 and 23) is present in the last SLFG, showing that the spiro atom 8 acts as the electronic connection of the two annular substructures.

Patulin (4). There are 4 FLFGs in the molecule. In this case the conjugated carbonyl system is found at the first level (cf. adrenosterone), emphasizing the different weight that the system can have in a highly functionalized molecule where the group is not divisible into two parts.

The SLFGs include the two main pieces of the molecule: (a) the dihydropyran ring plus the hemiacetal oxygen; and (b) the oxodihydrofuran ring plus the hemiacetal oxygen. The last SLFG comprehends the whole structure.

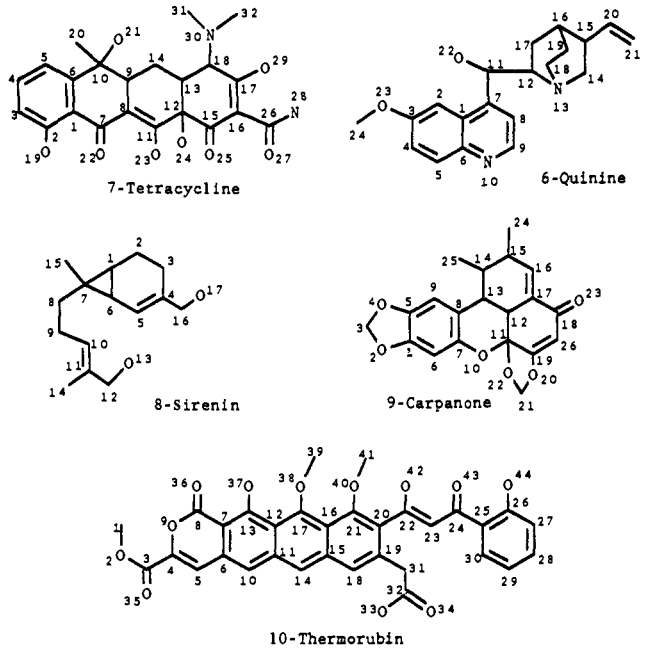

Figure 2. Atom numbering for structures **6–10**.

Table V. (De)stabilization Energies of FG Atoms[a]

molecule	atom	energy[b]	atom	energy[b]	atom	energy[b]
8	3	0.299	4	1.722	5	1.728
	6	0.313	9	0.307	10	1.729
	11	1.723	13	0.809	14	0.294
	16	0.199	17	0.809		
9	1	0.373	2	0.092	3	0.235
	4	0.092	5	0.375	6	0.281
	7	0.340	8	0.236	9	0.269
	10	0.106	11	0.104	12	0.300
	13	0.234	15	0.298	16	1.811
	17	1.544	18	5.070	19	1.805
	20	0.158	21	0.230	22	0.192
	23	5.701	26	1.449		
10	1	0.158	2	0.120	3	5.538
	4	1.936	5	1.833	6	0.720
	7	0.652	8	5.672	9	0.441
	10	0.744	11	1.071	12	1.016
	13	0.628	14	0.289	15	0.102
	16	0.068	17	0.393	18	0.934
	19	0.864	20	0.583	21	1.035
	22	1.637	23	1.543	24	5.341
	25	0.078	26	0.437	27	0.453
	28	0.424	29	0.403	30	0.408
	31	0.481	32	5.544	33	0.502
	34	5.455	35	5.736	36	5.811
	37	0.477	38	0.060	39	0.151
	40	0.075	41	0.151	42	0.503
	43	5.897	44	0.492		

[a] Atom numbers refer to Figure 2. [b] Energy is in a.u.

Prostaglandin E1 (5). There are 5 FLFGs in the molecule. They all correspond to chemical intuition. The two alcohols are found as separated entities at this level and they will be combined with the rest of the molecule in the next step, as shown by the SLFGs. These include (a) the cyclopentanone ring with the chain double bond, (b) the β-hydroxy ketone, and (c) the combination of each hydroxyl with the double bond.

The last SLFG is composed of the cyclopentanone, the hydroxyls, and the double bond, leaving the carboxyl apart.

Quinine (6). This molecule presents the only case in the set where an aromatic system is entirely part of one FLFG (1-2-3-4-5-6-7-8-9-10-11-12-23). The reason for this is the concentric disposition of the energies around the nitrogen atom that allows for a continuous FG growth.

Among the SLFGs there are 3 that are interesting due to saturated bond combinations including the tertiary amine; (a) with the aromatic ring; (b) with the vinyl; and (c) with the hydroxyl 11-22.

The last SLFG shows another case of a "long distance" relationship between apparently isolated FGs, combining the vinyl with the aromatic ring through the amine and the alcohol; the relation can be useful, e.g. in synthesis planning, showing a series of two FG relationships.

Tetracyclin (7). This molecule is another example of both the division of an aromatic ring and the identification at the first level of conjugated carbonyls. There are 4 conjugated carbonyls in the molecule: 1-6-7-22, 7-8-11-22, 15-16-17-25, and 16-17-26-28. They are identified by either one carbonyl (7-22) and two double bonds (1-6 and 8-11) or two carbonyls (15-25 and 26-28) and one double bond (16-17). Only the last two groups are identified as FLFGs, showing their "importance" in the structure.

There are many SLFGs: (a) alternative subdivisions of the aromatic ring; (b) β-dicarbonyls; (c) an α-amino alcohol; (d) "long distance" relations including β-dicarbonyls and double bonds, enols and hydroxyls, unsaturated carbonyls and hydroxyls or amines; (e) a long distance relation between the aromatic ring and the enol 8-11-23; and (f) the almost complete tetrahydronaphthalene (only atom 2 is excluded).

The last SLFG comprehends all the atoms in the molecule less the saturated and separated 13, 14, and 20 atoms.

Sirenin (8). This molecule presents only very simple FGs and the procedure furnishes the expected result. SLFGs are the two allylic alcohols.

Carpanone (9). There are 7 FLFGs that represent an exciting subdivision of the molecule in parts that are certainly uncommon FGs: e.g. atom 7 is not part of the phenyl ring; atom 13 is present in 4 FGs and therefore will function as the connection for the biggest part of the molecule; the first ketal (2-3-4) is a self-standing FG while the second (20-21-22) is not; and the two conjugated carbonyls (16-17-18-23 and 18-19-23-26) are not FLFGs, but they are part of two different SLFGs and also of the same last SLFG.

Among SLFGs there are (a) the complete aromatic ring, (b) combinations of the ketal part of the aromatic ring with the double bond 16-17 and the ketone 18-23, (c) combinations of the phenol part of the aromatic ring (atoms 7-10) with the same double bond and ketone, and (d) the two conjugated carbonyls.

The last SLFG comprehends all the FG atoms, atoms 21 and 22 excluded.

Thermorubin (10). This complex molecule represents a good challenge for the procedure and furnishes some interesting results. There are 14 FLFGs, the biggest one (10-11-12-13-14-15-17-37-38) being made up of only 9 atoms, a small number compared to the molecular size. The large number of highly important atoms (the molecule has 4 carbonyls, 4 aromatic rings, 8 alcoholic oxygens, 12 oxygens, and 28 unsaturated and only 4 saturated carbons) are evenly distributed, therefore causing the division of the molecule into pieces.

This molecule represents the only case with more than one combination of SLFGs. There are, in fact, 4 combined SLFGs: the first is the lactone ring; the second is the anthracene part; the third is the diketone part; and the fourth is the isolated phenyl ring. The carboxylic acid 31-32-33-34 also remains isolated (as an FLFG).

The first SLFG is a clear separation of the nonaromatic lactone from the aromatic part of the molecule and is a self-standing group. The second one is a more interesting case; in fact atoms 16, 20, and 21 are not in the group, even if they are part of the anthracene substructure, therefore showing the most important dependence of these atoms on the diketone part. The third group comprehends all the atoms in the diketone plus part of the anthracene (v.i.) and of the isolated phenyl (atom 25). Looking at the energies reported in Table V it is possible to note that atoms 16, 20, and 25 are not highly "important", while atoms 21, 22, and 24 are highly more "important" than all the near aromatic atoms; this is sufficient to separate the diketone and its neighbors from the molecule. (By changing the enolization sense of the diketone (toward atom 24) the result is even more striking, in fact atom 25 becomes a non-FG atom.) The last FG (31-32-33-34) is the only group separated

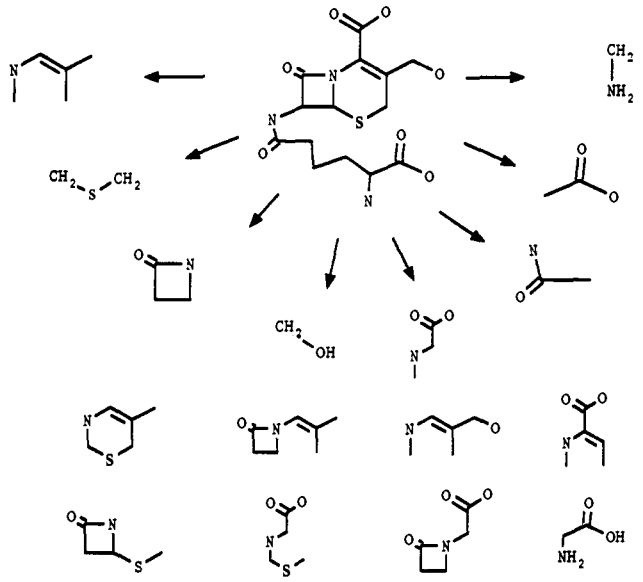

Figure 3. First and second level functional groups identified for cephalosporin-C. Arrows point to FLFGs.

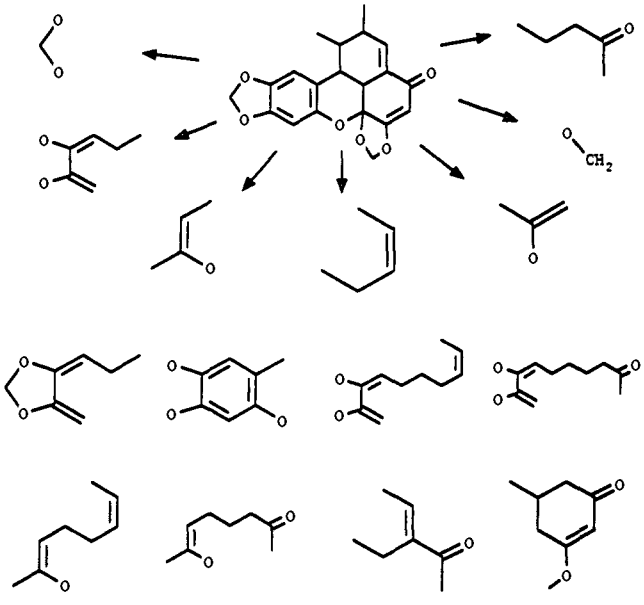

Figure 4. First and second level functional groups identified for carpanone. Arrows point to FLFGs.

by a saturated carbon from the molecule.

The FLFGs and the SLFGs for cephalosporin-C and for carpanone are also sketched in Figures 3 and 4; it is therefore possible to graphically observe the combination of FLFGs into SLFGs and the partitioning of the structures in functionalized substructures.

Eventually some comments on the complete set are required. The calculation of the (de)stabilization energies related to the presence of each atom gives a straightforward procedure for the "recognition" of the functional groups in all the molecules considered. It is possible to note a clear tendency of the procedure: the most "important" atoms (such as in carbonyls, double bonds) have the highest energies showing their potential reactivity; on the other hand, less reactive atoms (like in alcohols, amines) have low energies confirming the common opinion on their scarce "activity".

The complexity of the molecules is not a determining factor for very "important" atoms, while it can have great influence on other atoms whose energies are leveled by the possibility of distributing their interaction energies on many neighbors.

The found functional groups can be very different from the usual groups. In fact, the most common FGs are identified only in relatively simple molecules, where the separation between the groups is clear; in any other case the groups identified are correlated with the known ones only with difficulty. On the other hand, the FG definition reported above is univocal and self-consistent. It is therefore possible to find FGs in any molecule whatever its complexity. Moreover, the FGs are made up of atoms that are really interacting (their interactions are ensured by their contributions to the molecular energy); thus the reaction of one of them modifies the whole FG and is influenced by all the other FG atoms. This property is distinctive of an FG and it is much more useful than a rigid classification, because it can easily distinguish two similar FGs (e.g. a ketone and an aldehyde; an amide and an imide) and even find new FGs (e.g. double conjugated ketones).

Furthermore, in a deeper analysis of the results, it is possible to get some hints on the potential reactivity suggestions given by the present approach. Examining, for example, the reactivity of the carbonyls present in the molecule set, the quantitative differentiation among them is correlative with their reactivity. By simply taking the difference between the oxygen and carbon energies a correct ordering is obtained: acids > amides > ketones > unsaturated acids or esters > benzylic ketones > unsaturated ketones > unsaturated and benzylic ketones. The carbonyls that seem to deviate from this ordering are representative of particular situations; they are also in the right position: an unsaturated ketone with the double bond in common with another carbonyl and with an α-oxygen is less reactive than unsaturated ketones; a ketone with two α double bonds is even less reactive; a benzylic lactone is more reactive than an unsaturated ester.

This simple example shows the potential utility of the present approach that gives a definition of FG but also quantifies the importance of each FG atom.

Conclusion

Functional groups are of fundamental importance in organic chemistry where they are used both as molecular descriptors and as characteristics of the potential molecular interactions. Their univocal definition is still missing, and different kinds of difficulties seem to forbid a worldwide accepted formula.

A procedure for their identification has been reported together with a possible self-consistent definition based on the calculation of an atom descriptor and not on a qualitative "graphical" appearance. It is thus possible to find with certainty the presence of FGs in any molecule without defining a priori structural requirements.

The procedure applied to some examples has shown its applicability. The results are consistent and the molecular complexity is not a limiting factor.

Acknowledgment. Partial financial support by the "Consiglio Nazionale delle Ricerche" and by the "Ministero dell'Universita' e della Ricerca" is gratefully acknowledged.

Registry No. 1, 382-45-6; **2**, 61-24-5; **3**, 126-07-8; **4**, 149-29-1; **5**, 745-65-3; **6**, 130-95-0; **7**, 60-54-8; **8**, 19888-27-8; **9**, 26430-30-8; **10**, 37577-75-6.

Supplementary Material Available: Derivation of the molecular electron energy (4 pages). Ordering information is given on any current masthead page.