match no. 24 pp. 39-49 1989

MATHEMATICAL DESCRIPTION AND ENUMERATION OF SIMPLE CHEMICAL FUNCTIONAL GROUPS

René BARONE and Michel CHANON

Laboratoire de Chimie Inorganique Moléculaire. Case 561.
UA CNRS 126 . Faculté des Sciences de St Jérôme.
13397 Marseille cedex 13. France.

(Received: May 1988)

ABSTRACT

A new system is presented for mathematical description of chemical functional groups. The system is an extension of the Hendrickson 's (6) . We developed a computer program using this system to calculate all the possible interconversions between the groups, to display these interconversions on the screen and to find the shortest pathways linking two groups.

INTRODUCTION

One of the most fascinating challenges in the field of chemical information and computerized chemistry is to find a simple mathematical description of compounds in order to manipulate chemical knowledge simply and quickly, by means of simple mathematical calculations.

This need is very important in the field of computer-aided organic synthesis. In this field compounds are generally described by atom-bond tables (1), by a BE matrix in the EROS program (2), by the linear DARC system in the SYNOPSYS program (3) and by the character of carbon atoms in the SYNGEN program (4). Kratochvil and coworkers proposed also a simple description of atoms (5) and Camps (6) proposed an extension to Hendrickson's system .

From these descriptions, only the models of Hendrickson and Kratochvil allow one to code the reactions by a simple addition or subtraction, the other programs needing more sophisticated computer programs.

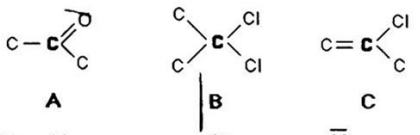
But these two models do not allow one to describe all the functional groups with accuracy , so we searched for a more detailed model.

We present in this paper a new description of simple chemical functional groups in order to develop a new computer-aided organic synthesis program. We present the first step of this project : the construction of all the theoretically possible interconversions between functional groups.

MATHEMATICAL MODEL

The model that we have developed presents some analogies with those of Hendrickson (4a) and Kratochvil (5) in that we describe only carbon atoms and their first environment.

The Hendrickson model is very general to describe a class of functional groups by a number but does not distinguish all the functions, for example the atoms A and B of scheme 1 have the same character 22 : 2 σ bonds and 2 "functional" bonds.



	A	B	C
Hendrickson :	22	22	$\overline{22}$
Kratochvil :	0210	0400	0210

Scheme 1

In the Kratochvil model (5) , only the number of free electrons, single, double and triple bonds are given : atoms A and C, of scheme 1, are described by the same formula : 0210 (0 free electron, 2 single bonds, 1 double bond and 0 triple bond).

In Hendrickson's model four kinds of attachment are considered : H (bond between carbon and hydrogen) , R (σ bond between two carbons) , Π (π bond between two carbons) and Z (σ and π bonds between carbon and heteroatoms). A

transformation at a carbon is described by two letters, the first one indicates the bond formed and the second one the bond broken. For example in the reaction :

Cl -H + C = C -> Cl - C - C - H
1 2

For carbon 1 the kind of reaction is ZΠ (formation of a C-Z bond, and loss of a Π CC bond) and for carbon 2 the code is H Π (+H , - Π). From these 4 attachments there are 16 possible reactions at a carbon site (4a).

In order to describe all the possible functional groups we searched for a more detailed model. In a first attempt we developed a code with four digits, 2 for the CC bonds and 2 for the C-Heteroatom bonds (designated by CZ in this article), and for each bond we distinguish the σ and π bonds. The code was :

$C = 1000 * \sigma CC + 100 * \pi CC + 10 * \sigma CZ + \pi CZ$

and for the carbon atoms of scheme 1 we obtain : A = 2011, B = 2020 and C = 2110.

One may remark that the sum of the 3 last digits gives the f value of the Hendrickson code (4a) and so it would be possible to generalize the class of the functional group, if needed .

All the possible groups may be easily written by hand or calculated by a simple program. They are displayed in column 1 of table 1.

With this model we are considering five attachments : H , σCC , πCC, σCZ and πCZ, where σCC and πCC are identical respectively to the R and Π values of Hendrickson, whereas σCZ and πCZ represent σ and π bonds between carbon and heteroatoms. Having defined these five attachments, a carbon atom may then undergo 25 formal reactions which are listed in table 2. The bond made (+) and the bond broken

No		model 1	model 2	No		model 1	model 2
1	C	00 00	000 000	22	C-C-Z	10 10	100 100
2	C-C	10 00	100 000	23		20 10	200 100
3	C-C-C	20 00	200 000				
4		30 00	300 000	24		30 10	300 100
5		40 00	400 000	25	Z-C=C	11 10	010 100
6	C=C	11 00	010 000	26		21 10	110 100
7	C-C=C	21 00	110 000				
8		31 00	210 000	27	Z-C≡C	12 10	001 100
9	C=C=C	22 00	020 000	28		10 20	100 200
10	C≡C	12 00	001 000	29		20 20	200 200
11	C-C≡C	22 00	101 000				
12	Z-C	00 10	000 100	30		11 20	010 200
13	Z-C-Z	00 20	000 200				
14		00 30	000 300	31		10 30	100 300
15		00 40	000 400	32	C-C=Z	10 11	100 010
16	C=Z	00 11	000 010	33		20 11	200 010
17	Z-C=Z	00 21	000 110				
18		00 31	000 210	34	C=C=Z	11 11	010 010
19	Z=C=Z	00 22	000 020	35		10 21	100 110
20	Z≡C	00 12	000 001				
21	Z≡C-Z	00 22	000 101	36	C-C≡Z	10 12	100 001

Table 1. Codes developed to describe simples chemical functions.

(-) are indicated for each case. For the third case (+ H - πcc) we show only the modification on one carbon and we do not specify what is added on the other carbon of the double bond. This model presents nine more reactions than Hendrickson's (4a).

+	-	Examples
H	H	**C**-H → H-**C**
H	σcc	C-**C** → H-**C**
H	πcc	**C**=C → H-**C**-C
H	σcz	**C**-Z → H-**C**
H	πcz	**C**=Z → H-**C**-Z
σcc	H	**C**-H → C-**C**
σcc	σcc	C-**C** → **C**-C
σcc	πcc	**C**=C → C-**C**-C
σcc	σcz	**C**-Z → C-**C**
σcc	πcz	**C**=Z → C-**C**-Z
πcc	H	C-**C**-H → C=**C**
πcc	σcc	C-**C**-C → C=**C**
πcc	πcc	C-**C**=C → C=**C**-C
πcc	σcz	C-**C**-Z → C=**C**
πcc	πcz	C-**C**=Z → C=**C**-Z
σcz	H	**C**-H → Z-**C**
σcz	σcc	**C**-C → Z-**C**
σcz	πcc	**C**=C → Z-**C**-C
σcz	σcz	**C**-Z → Z-**C**
σcz	πcz	**C**=Z → Z-**C**-Z
πcz	H	Z-**C**-H → Z=**C**
πcz	σcc	Z-**C**-C → Z=**C**
πcz	πcc	Z-**C**=C → Z=**C**-C
πcz	σcz	Z-**C**-Z → Z=**C**
πcz	πcz	Z-**C**=Z → Z=**C**-Z

Table 2 . Types of reactions at one carbon site

However, in spite of its qualities, this model does not distinguish the functions 9 and 11 and 19 and 21 of table 1 (scheme 2) :

C - **C**≡ C and C = **C**= C code = 22 00

N - **C**≡ N and N= **C**= N code = 00 22

scheme 2

So we developed another model indicating clearly the number of single, double and triple bonds for the CC and C-heteroatom bonds. We obtain a code with 6 digits, the first three for the CC bonds and the others for the C-heteratom bonds :

C = C1 C2 C3 Z1 Z2 Z3

in which Ci designates the number of CC bonds with multiplicity i, i.e. the first three digits indicate the number of single, double or triple CC bonds ; similarly for Zi, i.e. the last three digits denote the number of single, double or triple C-heteroatom bonds. Thus the ambiguities indicated above are removed (scheme 3) :

C - **C**≡ C code = 101 000 N - **C**≡ N code = 000 101

C = **C** = C code = 020 000 N = **C** = N code = 000 020

Scheme 3

Column 2 of table 1 lists the code obtained for all the functional groups.

With this model the value of f (Hendrickson's model) is given by the formula :

f = C2 + 2 * C3 + Z1 + 2 * Z2 + 3 * Z3

COMPUTER PROGRAM

With the developed codes we constructed an functional group interconversion (FGI) matrix between the groups. The elements of the matrix are : FGI (I,J) = K in which K = 1 if there is a reaction in one step from group number I to group number J, and K = 0 if there is no direct reaction between the two groups.

Then from this interconversion matrix we developed a subroutine which allows us to construct a pathway between any two groups .

The construction of such a matrix is done by a BASIC program which has the following steps :

FOR I = 1 TO N -1 (N = number of groups)

FOR J = I + 1 TO N

is there a one step reaction between groups I and J ?

if yes then FGI (I,J) = 1 and FGI (J,I) = 1

NEXT J

NEXT I

To construct such a matrix we followed some rules which predict if a reaction in one step is theoretically possible between two groups. For each reaction the program calculates the variation Δ of the number of CC and C-heteroatom bonds, and the variation of sigma and π bonds (CC and CZ); if one of these $|\Delta| > 1$ then the reaction is discarded. If the number of variations is greater than 2 the reaction is also discarded : it means that several elementary reactions occurred.

Some examples are given in table 3 :

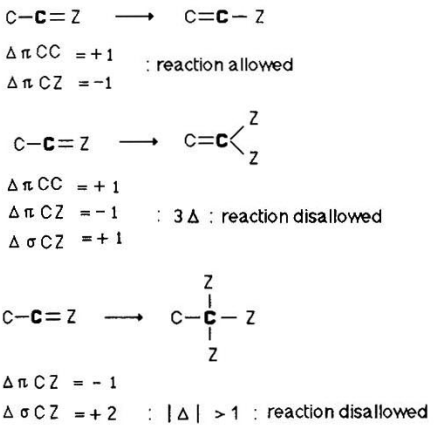


Table 3. Examples of evaluation made by the program.

These calculations are easily made by means of the descriptions adopted and the matrix obtained is presented in figure 1.

This matrix indicates the theoretical possible interconversions for each group. For example, the elements (2,1) , (2,3) , (2,6) , (2,12) and (2,22) are equal to 1 , which indicates that a one step reaction is theoretically possible between group 2 and groups 1, 3 , 6 , 12 and 22 (scheme 4) :

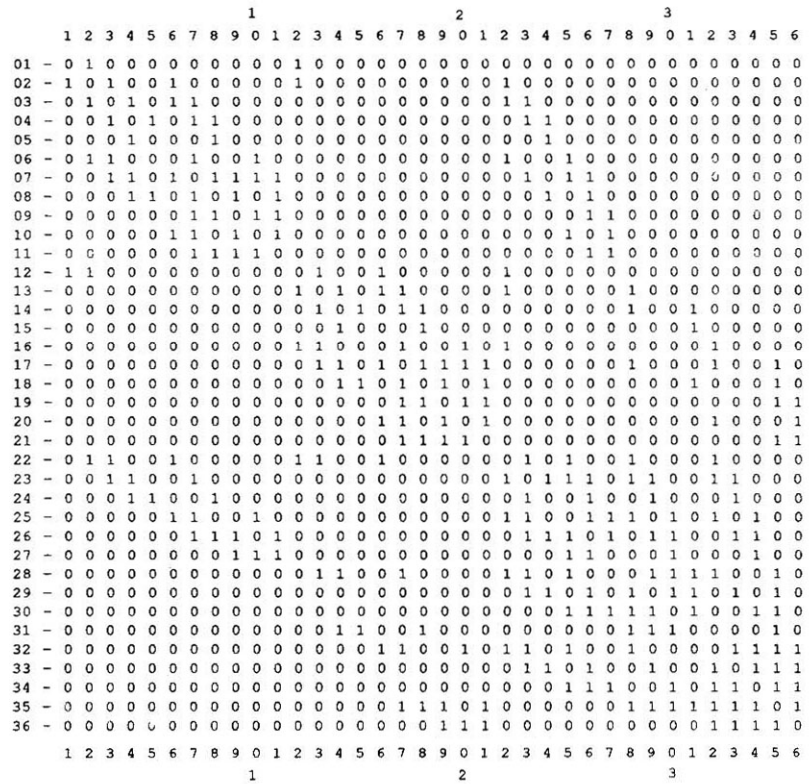
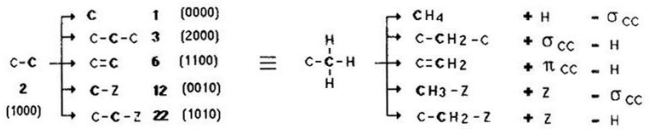


Figure 1 . Interconversion matrix for the 36 groups



Scheme 4

The program offers two options : the first one is the visualization of the interconversions and the second one is the construction of a pathway between two groups.

In option 1, the screen of the microcomputer is divided into two parts and in each window the user may display the different groups. He has again two possibilities : (1i) he may visualize what it is possible to obtain from a group : in this case the user selects a group in window A and has the other groups defile i n window B ; (1ii) he may visualize the precursors of a group : in this case the group of window B is fixed and the user has the precursors defile in window A. Figure 2 shows the screen of the computer when option 1 is selected.

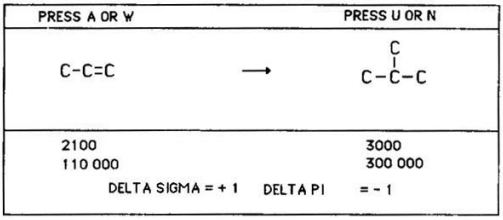
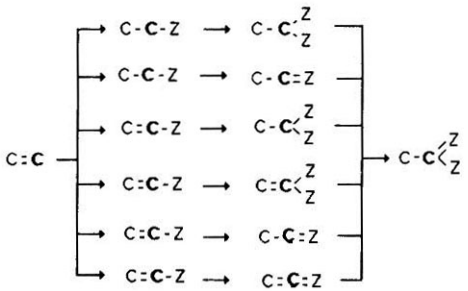


Figure 2. A view of the screen when the user lists the possible interconversions. An option allows the user to display left and right groups one by one. For each reaction presented, the result of the calculations for its validity is indicated.

When option 2 is selected the user has to choose two groups. This is done by having the groups defiling in each window. Then the program searches the shortest pathways between them and displays the theoretical reactions. Figure 3 shows some interconversions found by the program.



If we display hydrogens we obtain :

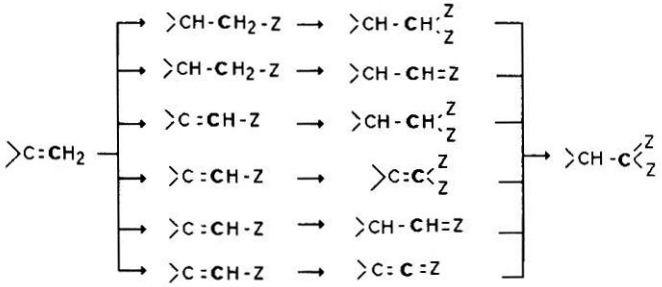
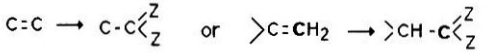


Figure 3 . Solutions given by the program for the problem :



CONCLUSION

We propose a new code to describe simple chemical functions. From this code by simple mathematical calculations we developed an interconversion matrix between the different functions and we developed a computer program which allows one to find a pathway between two groups. Work is under development to describe more complicated functional groups.

The program is written in BASIC and runs on an IBM/PC. It is available from the authors on request.

Acknowledgement : We thank Prof. J.B. Hendrickson for useful suggestions during the writing of this paper.

REFERENCES

1 - R. BARONE and M. CHANON, "Computer-aided Organic Synthesis" in "Computer Aids to Chemistry", G. VERNIN and M. CHANON Eds. E. Horwood Ltd, Chichester, 1986.

2 - I. UGI, J. BAUER, J. BRANDT, J. FRIEDRICH, J. GASTEIGER, C. JOCHUMS, W. SCHUBERT, Ang. Chem. Int. Ed., 1979, **18**, 111.

3 - R. PICCHIOTTINO, G. GEORGOULIS, G. SICOURI, A. PANAYE, J.E. DUBOIS, J. Inf. Chem. Inf. Comput. Sci., 1984, **24**, 241.

4- (a) J. B. HENDRICKSON, J. Am. Chem. Soc., 1971, **93**, 6847.
(b) J. B. HENDRICKSON, D. L. GRIER, A. G. TOCZKO, J. Am. Chem. Soc., 1985, **107**, 5228.

5 - J. KOCA, M. KRATOCHVIL, M. KUNZ, V. KVASNICKA, Coll. Czech. Chem. Commun., 1984, **49**, 1247.

6 - R. GOVIND and G.J. POWERS, in "Computer-Assisted Organic Synthesis ", W.T. Wipke and W.J. Howe, Eds., ACS Symp. Ser. **61**, 1977, page 81, ref. 3.