

## Microcomputer and Organic Synthesis. 3. The MARSEIL/SOS Expert System, a New Graphic Approach. An Electronic Lab Note for Organic Synthesis

Patrick Azario, René Barone,\* and Michel Chanon

Laboratoire de Chimie Inorganique Moléculaire, UA CNRS 126, Faculté des Sciences de St. Jérôme, Avenue Escadrille Normandie-Niemen, 13397 Marseille Cedex 13, France

Received April 3, 1987

MARSEIL/SOS is a new computer-aided organic synthesis program. We describe its main characteristics: (i) the program is open to the user who may add his own perception of the target; (ii) the input of reactions, i.e., the knowledge base of the system, is mainly graphic; (iii) a graphical description of the bibliography is associated with each reaction; (iv) the possibility of graphical evaluation of solutions; and (v) the possibility to modify the evaluation tests during the retrosynthesis, which gives a property of self-learning to the program. These features allow a quick and easy way to enter new reactions. New syntheses for yohimbine and bulnesol are presented. MARSEIL/SOS runs on a Macintosh+ microcomputer and is user-friendly thanks to the mouse, scrolling menus, and multiple windows.

Computer-aided organic synthesis (CAOS) is now almost 20 years old<sup>1</sup> and has evolved in divergent directions<sup>2</sup> since the first days. Originally mainly backward, it has been more recently developed in the forward direction to give the high-performance CAMEO<sup>3</sup> or, in the specific field of catalysis induced by transition-metal complexes to give TAMREAC.<sup>4</sup> On the other hand, the relative positions of strategies and tactics have been clarified with the simplifying and organizing approaches of Hendrickson.<sup>5</sup> Unfortunately, except in some famed industrial firms, the computer approach has not yet very much changed the everyday life of the synthetic chemist. This partly comes from the cost of the programs and hardware equipment. To circumvent these drawbacks and to take advantage of the dramatic progress in microcomputer capabilities, we have been developing for more than 5 years an approach<sup>6</sup> able to be used in any laboratory. Before describing it, we must clearly state that it is not competitive with sophisticated approaches but complements them in the same way as HMO methods complement ab initio ones in molecular orbital theories.

The main characteristics of MARSEIL/SOS are that (1) it works on a Macintosh, an inexpensive, widely available microcomputer, (2) it has a starting basis of 350 reactions covering approximately the content of the third edition by March<sup>7</sup> (this basis may be easily extended by the user to adapt the program to his specific field), (3) the evaluation of reaction feasibility is self-improving as it starts from a basis but may also be continuously and directly improved by the user, and (4) the design of MARSEIL/SOS, being centered as it is on the simple, everyday graphical representation of structures, reactions, and evaluation, makes it friendly to even a non computer specialist. The aim of this paper is to detail the foregoing points that establish MARSEIL/SOS as the first program usable as an electronic lab note for organic synthesis, extending the field of computer aided synthesis to every

medium-sized synthetic laboratory.

### Description of MARSEIL/SOS

MARSEIL/SOS is the first program of a general system called MARSEIL (Multi Approaches for the Research of Synthesis by Efficient and Interactive Logic). This system will offer several programs to help the chemist in solving a synthetic problem; SOS (Simulated Organic Synthesis) is the CAOS program of MARSEIL. Other programs have been developed or are still in development, such as TAMREAC<sup>4</sup> and REKEST (REsearch for the KEy STep),<sup>8</sup> which helps the chemist in finding the general strategy to reach a given target.

MARSEIL/SOS follows the retrosynthetic approach previously developed for SOS.<sup>9</sup> The flow chart of Figure 1 summarizes the main steps of the program.

**A. Input of the Target.** The chemist/computer communication is made possible by the mouse and the scrolling menus. The chemist moves the mouse on the desk and a cursor follows the movement on the screen. When he clicks on the button of the mouse, an atom is drawn on the screen and added in the table of atoms that describes the compound; a bond is created with the previously entered atom (excepted, of course, for the first one) and is added in the table of bonds. By clicking twice on the last atom, the mouse is made free, and one may come back to a previous atom in order to enter a substituent. The choice of atoms, bonds, and charges is made by means of pull-down menus. Figure 2 shows the content of the atoms menu. In the menu, L stands for a leaving group, X for a halogen, T for any heteroatom, Nu for a nucleophilic center, and R for a carbon chain. The option "others" allows one to choose an atom in the periodic table or a superatom such as Z for a withdrawing group, Ar for an aromatic ring, or A for any atom. The program may handle compounds with up to 64 atoms (other than H). The atoms and bonds tables that describe the target are similar to the ones used in the SECS<sup>10</sup> program.

**B. Perception.** When the input of the target is done, the connectivity table is developed and the program analyzes it in order to extract a binary description in which the main features of the target (rings, functions, nucleophilic centers, etc.)<sup>9</sup> are coded. The perception of structural features is one of the most important parts of any

(1) Corey, E. J.; Wipke, W. T. *Science (Washington, S.C.)* 1969, 166, 178.

(2) (a) Barone, R.; Chanon, M. *Computer Aids to Chemistry*; Vernin, G., Chanon, M., Eds.; E. Horwood: Chichester, 1986. (b) Bersohn, M.; Esack, A. *Chem. Rev.* 1976, 76, 269. (c) Long, A. K.; Rubenstein, S. D.; Joncas, L. *J. Chem. Eng. News* 1983, 61, 22. (d) Haggin, J. *Chem. Eng. News* 1983, 61, 7. (e) Gund, P. *Annu. Rep. Med. Chem.* 1977, 12, 288.

(3) Gushurst, A. J.; Jorgensen, W. L. *J. Org. Chem.* 1986, 51, 3513.

(4) Theodosiou, I.; Barone, R.; Chanon, M. *Adv. Organomet. Chem.* 1986, 26, 165.

(5) Hendrickson, J. B. *Acc. Chem. Res.* 1986, 19, 274.

(6) Barone, R.; Chanon, M.; Cense, J. M.; Cadiot, P. *Bull. Soc. Chim. Belg.* 1982, 91, 333.

(7) March, J. *Advanced Organic Chemistry*, 3rd ed.; Wiley: New York, 1986.

(8) Barone, R.; Chanon, M. *Chimia* 1986, 40, 436.

(9) Barone, R.; Chanon, M. *Nouv. J. Chim.* 1978, 2, 659.

(10) Wipke, W. T. *Computer Representation and Manipulation of Chemical Information*; Wipke, W. T., Heller, S. R., Feldmann, R. J., Hyde, E., Eds.; Wiley-Interscience: New York, 1974, p 147.

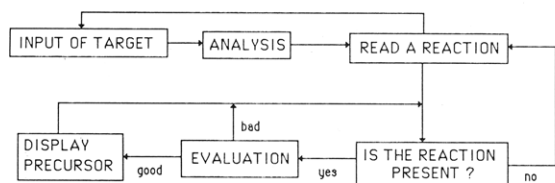


Figure 1. Main steps of the program.

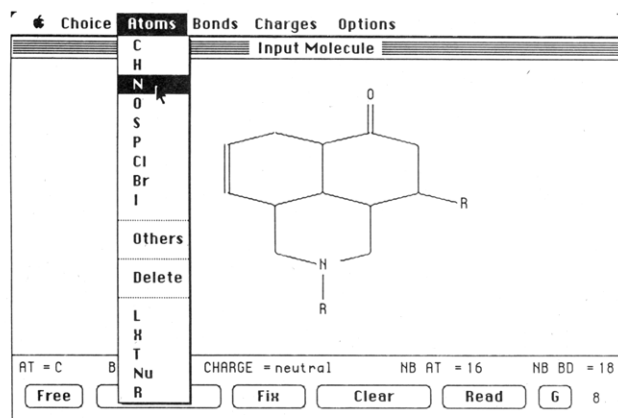


Figure 2. Input of target. Scrolling menus allow the selection of options.

CAOS program because it is used in the search of reactions and their evaluation.

In all other programs, this part is done by one or several subroutines, which are modifiable only by the writers of the program. The users have no possibility to modify this perception.

We developed a new approach, which allows the user to introduce his own perception at will. Thus, he has the possibility to introduce the particular functions of his specific field of chemistry.

The perception of the structural features of the target is done in two part. The first one is a "classically" programmed one: routines have been written to find rings and standard features such as nucleophilic centers, withdrawing groups, etc.; 60 groups are found by this approach. The second part is made from a file of substructures, which are matched with the target. These substructures constitute a modifiable knowledge base of the target. The input of these substructures is done by drawing them on the screen as done for the target. A graphical editor is provided to the user to enter and/or modify this data base, so he may add the desired groups. The program may handle up to 512 groups. The time to analyze the target is proportional to the number of substructures to find. For example, the perception of the 60 groups and 25 "graphical" substructures in a target with 15 atoms and 3 rings is made in 20 s. This method is slower than the programmed one because the program has to make an atom by atom match from the tables that describe the target and the substructures. We prefer it, however, because the program is open to the user who may add the desired substructures, therefore adapting the program to his own chemistry.

**C. Description of Transforms.** The search for a transform is done in three parts: (i) looking for a characteristic substructure, (ii) evaluation, and (iii) building of the precursor. In the first version of SOS, the input of all these different data was alphanumeric and, therefore, tedious. In contrast, the graphical possibilities of Macintosh allow the input of a transform in a natural way, i.e., graphically, by drawing it on the screen, which is divided into three windows, one for each part.

**C.1. Input of the Substructure That Describes the Transform.** This is drawn by the user in the substructure

Scheme I

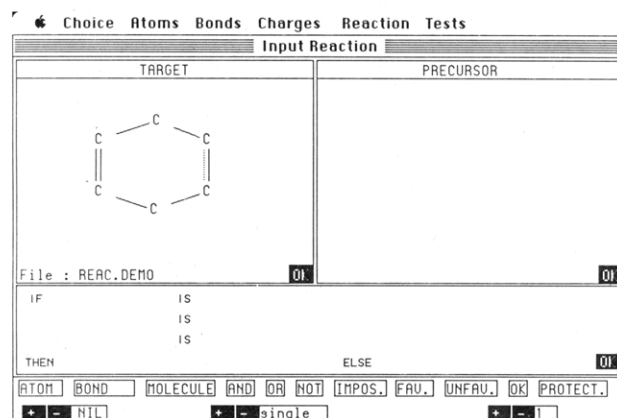
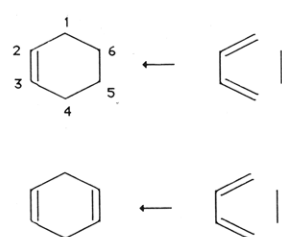


Figure 3. Input of the substructure, which describes the reaction. The screen is divided in three parts, the first one for the input of the target substructure, the second one for the precursor, and the third one for the alphanumerical tests.

window as for the input of the target but with more possibilities. For example, in order to generalize some reactions, one indicates that a bond may be single or double, or double or triple, and such bonds may be visualized on the screen. In the Diels-Alder reaction, the bond between atoms 5 and 6 may be single or double (Scheme I), and one describes these two cases by only one reaction (see Figure 3). The substructures are internally described by an atom-bond table similar to the one developed for the target. They are found in the target according to an atom by atom matching of the connectivity tables and through the binary description of the target.<sup>9</sup> Figure 4 shows the internal description of a substructure.

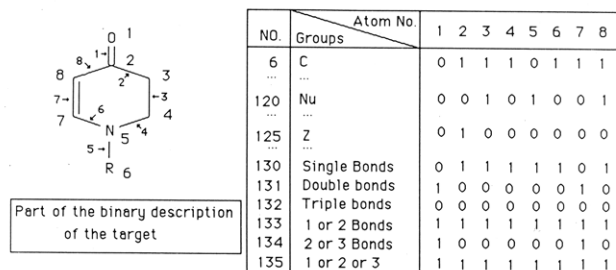
**C.2. Building of the Precursor.** The modifications to be brought to the substructure must be indicated in order to build the precursor. When the user clicks in the OK rectangle of the substructure window, the substructure is drawn in the second window, and the user modifies it. He may add, delete, or modify atoms and/or bonds. These changes are saved and constitute the set of modifications to be applied to the target to build the precursor (Figure 5).

**C.3. Evaluation.** This evaluation is performed by several tests, which check if some specific substructures susceptible of having an influence on the planned reaction are present in the target.

In some programs (LHASA, SECS, PASCOP, PSYCHO),<sup>2a</sup> the writing of these tests comes to the writing of subroutines in a special language: ChmTrn in LHASA,<sup>11</sup> Alchem in SECS,<sup>12</sup> Class in PSYCHO.<sup>13</sup> The writing of

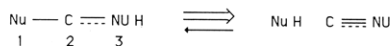
(11) (a) Pensak, D. A.; Corey, E. J. *Computer-Assisted Organic Synthesis*; Wipke, W. T., Howe, W. J., Eds.; ACS Symposium Series 61; American Chemical Society: Washington, D.C., 1977; p 1. (b) Corey, E. J.; Long, A. K.; Rubenstein, S. D. *Science (Washington, D.C.)*, 1985 228, 408.

(12) Wipke, W. T.; Braun, H.; Smith, G.; Choplin, F.; Sieber, W.; In *Computer-Assisted Organic Synthesis*; Wipke, W. T., Howe, W. J., Eds.; ACS Symposium Series 61; American Chemical Society: Washington, D.C., 1977, p 97.



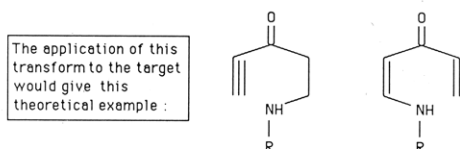
## DESCRIPTION OF TRANSFORM :

Substructure which describes nucleophilic addition :

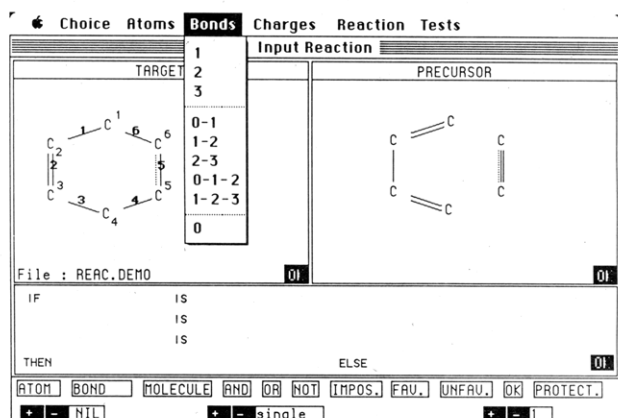


The internal description is :

Atom	Code	No. H	Bond	Atom1	Atom2	Order
1	120	0	1	1	2	130
2	6	0	2	2	3	133
3	120	1				



**Figure 4.** Binary description of the target and example of the description of a transform.



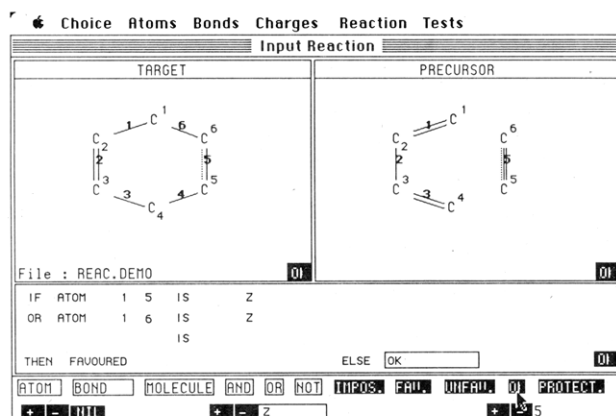
**Figure 5.** Input of the substructure precursor. Bond 5 (between atoms 5 and 6) may be single or double in the target and becomes double or triple, respectively, in the precursor. These bonds can be visualized on the drawing. The "bonds" menu shows the possible bond orders available.

such subroutines may be long and delicate.<sup>11a</sup>

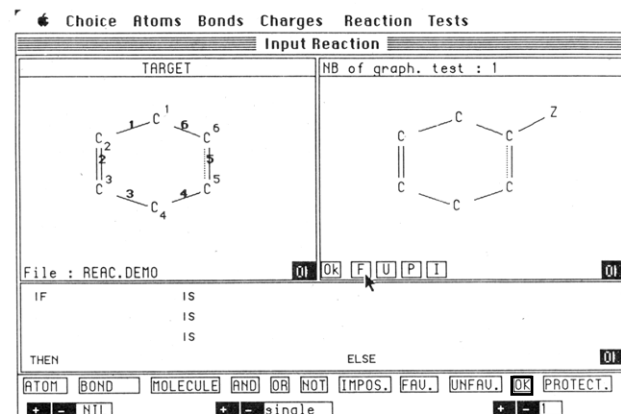
Our mechanistic approach, being more general, does not necessitate the development of a detailed description for the reactions. Therefore, it does not require the development of sophisticated tests. So, we prefer to develop simpler tests, without any links between them, without instructions such as GOTO, in order to input the knowledge without having to worry about the sequence. Each time a new test is added, it is stored at the end of the file. These tests have the following form, where Fi is a substructure, a group found during the analysis of the target (See part B)

```
IF          ATOM/BOND No. IS (NOT) F1 AND/OR (NOT) F2 AND/OR (NOT) F3
AND/OR ATOM/BOND No. IS (NOT) F4 AND/OR (NOT) F5 AND/OR (NOT) F6
AND/OR ATOM/BOND No. IS (NOT) F7 AND/OR (NOT) F8 AND/OR (NOT) F9
THEN ACTION1 ELSE ACTION2
```

ACTION may be: OK, impossible, favored, unfavored, or protect.



**Figure 6.** Input of tests. The possible keywords are activated (dark area): impossible, favored, unfavored, OK, protect. The test is: if there is an electron withdrawing group (Z) in  $\alpha$  (1) of atom 5 or 6, then the reaction is favored.



**Figure 7.** Input of a graphical test. If there is a withdrawing group (Z), then the reaction is favored.

The user need not learn the particular syntax necessary to write these tests: since they necessitate only a few words, the writing of these tests may be guided by the program, which activates the keywords to be used when required. The user just moves the cursor and clicks on the selected keywords. Figure 6 shows the tests window.

In order to complete this important part of the program, we have developed a new evaluation based upon graphical tests. Instead of describing a particular situation by means of sentences, it is possible to describe it by a drawing. The user activates this option in the menu and a new window appears where he draws a substructure and indicates an action: OK, impossible, favored, unfavored, or protect (Figure 7). If this substructure is present in the target, the action indicated will be executed. This option allows one to input complex tests, which would be difficult or impossible to enter by text.

One may add references with a graphical description of the reaction, by selecting the option BIBLIO. The user may then draw the full reaction and type the reference (Figure 8).

The reactions are stored in different files. The user may create the files that he wants. There are several files: one for the main mechanisms, one for rearrangements, and others files for the main functional groups (alkenes, acids, ketones, etc.).

In the present version, the number of atoms in a transform is limited to 12. It is possible to have 10 alphanumeric tests and 10 graphical tests. It is possible to store 200 transforms on a 800 Kb disk, but we plan to modify the structure of the data in order to have any number of tests for one reaction.

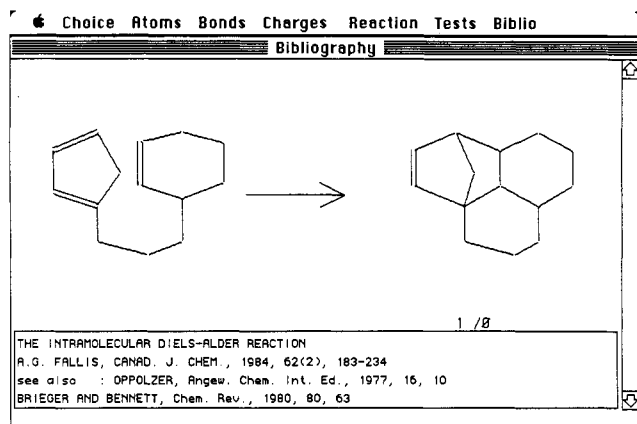


Figure 8. Input of references.

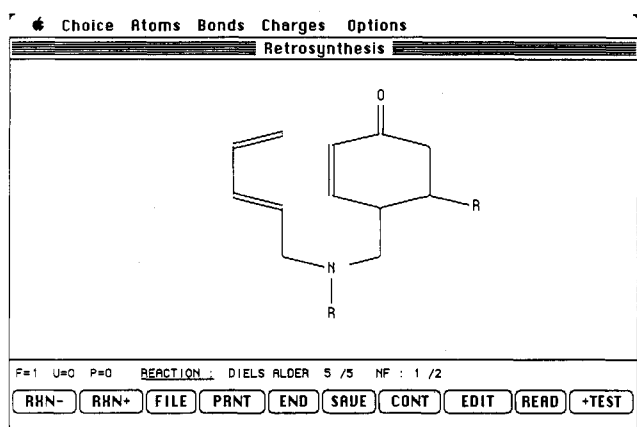


Figure 9. Output of the program. The main options are displayed at the bottom of the screen. Others options are available in the scrolling menus.

**D. Program Flow.** The flowchart of the program is similar to SOS.<sup>6,9</sup> Precursors are presented one after the other. Figure 9 shows the screen during this part: The user sees the solution and has several options at his disposal:

RXN-: returns to the preceding reaction.

RXN+: stops the search for the current reaction and jumps to the following one.

FILE: changes the current file. Reactions are saved in several files: mechanisms, syntheses of alkenes, acids, etc.

PRNT: hard copy of the screen.

EDIT: to move atoms in order to make a better drawing. It also allows changing the nature of an atom, for example the letter L represents a leaving group; the user may change L to Br, Cl, or whatever he wants.

Fx: there are x favorable conditions in the solution shown.

Uy: there are y unfavorable conditions in the solution.

Pz: there are z functions to protect.

To locate the different groups (favorable, unfavorable, protect), the user clicks on the letter, and the groups are displayed. Figure 10 illustrates this point.

If the reaction that is displayed is not familiar to the user, he may choose the option BIBLIO in the menu, and references for the reaction appear on the screen. Several references may be associated with one reaction.

If the solution presented is not good, because a test is not present, the user has the possibility to activate the test option (+TEST) in order to enter a new test for the current reaction. Therefore, the program has a kind of self-learning ability. Since the tests are unordered, it is possible to have such an option. The new test is added at the end of the file. When the chemist codes a reaction, it is difficult

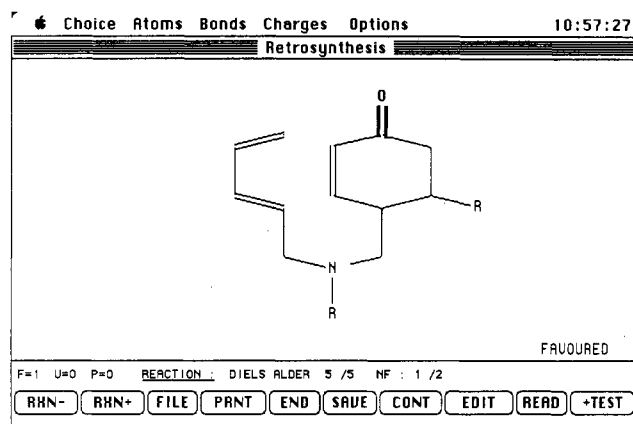


Figure 10. It is possible to display the group that favors the reaction.

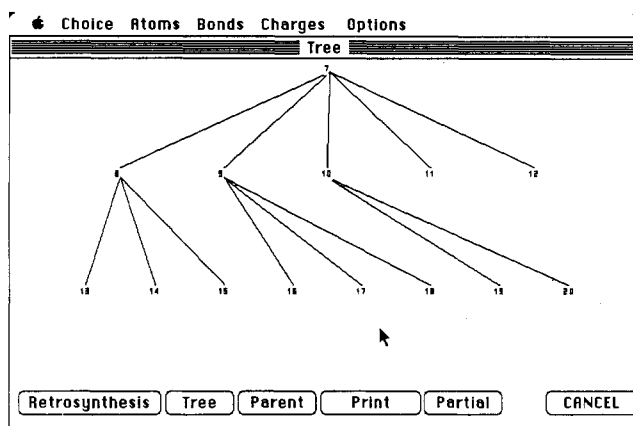


Figure 11. Retrosynthesis tree.

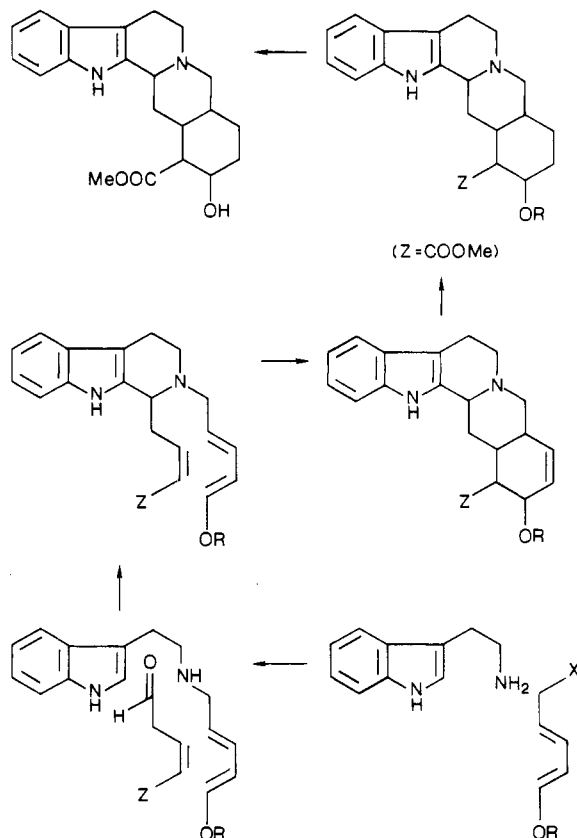


Figure 12. New synthesis of yohimbine found by the program. to foresee all the conditions influencing it. It is during the retrosynthesis that one can see the lacking tests. So, the

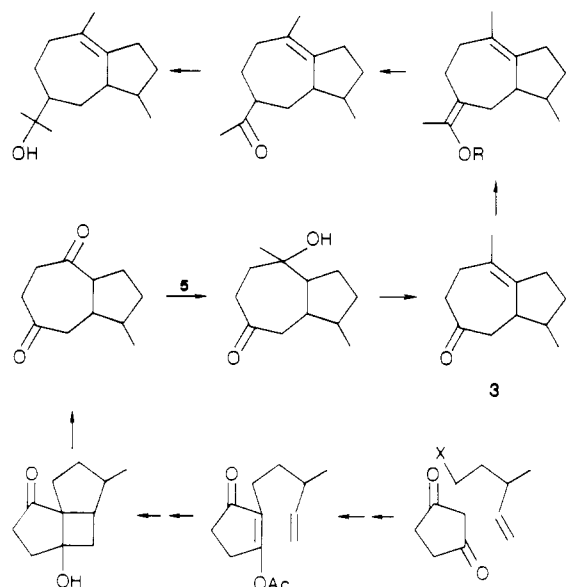


Figure 13. New synthesis of bulnesol found by the program.

program allows the simple addition of tests during the retrosynthesis and the immediate perception of their effects. This interactive procedure avoids time consuming shuttles: quite the retrosynthesis, enter in the input reaction module, come back to the retrosynthesis.

In order to visualize the retrosynthesis tree and/or select a new target among the precursors, the tree option is available. Figure 11 shows the screen when this option is activated. Precursors are stored on the internal drive of the Macintosh. The size of this file is 200 Kb. The number of saved precursors is a function of the number of atoms: approximately 300 precursors of 64 atoms or 600 of 32 atoms.

**E. Results.** MARSEIL/SOS has been used to solve several problems and is used by students.<sup>14</sup> Figures 12

and 13 show two new examples of synthesis found by the program for yohimbine<sup>15</sup> and bulnesol.<sup>16</sup> These schemes show the main weakness of the program: presently, there is no treatment of stereochemistry and this evaluation is left to the chemist. Nevertheless, the program is able to propose interesting ideas such as an internal Diels-Alder reaction to build the D and E rings of the yohimbine skeleton or an intramolecular De Mayo reaction to build the seven-membered ring of bulnesol. For step 5 in this synthesis, the program indicates that the other ketone must be protected. For the synthesis of bulnesol from precursor 3, the program proposes another sequence:  $\text{CO} \rightarrow \text{CHOH} \rightarrow \text{CHBr} \rightarrow \text{CHMgBR} + \text{CH}_3\text{COCH}_3 \rightarrow \text{bulnesol}$ .

### Conclusion

The program MARSEIL/SOS offers new features in the field of computerized organic synthesis design, such as personalized analysis of the target, which allows the user to easily add his own chemistry; graphical description of reactions, which allows their easy input; graphical description of bibliography; and the possibility of graphic evaluation. The program is also able to show the different groups that may influence the reaction. Then, because the program is interactive, it is also possible to add new tests during the retrosynthesis.

The next stage of this project will be the development of stereochemistry.

MARSEIL/SOS is written with MacFORTH and runs on a Macintosh+ microcomputer with one external drive.

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

(14) Bertrand, M. P.; Monti, H.; Barone, R. *J. Chem. Educ.* 1986, 63, 624.

(15) Martin, S. F.; Rueger, H. *Tetrahedron Lett.* 1985, 26, 5227 and references therein.

(16) Marshall, J. A.; Partridge, J. J., *Tetrahedron* 1969, 25, 2159 and references therein.

## Separation of Mass Law and Solvent Effects in Kinetics of Solvolyses of *p*-Nitrobenzoyl Chloride in Aqueous Binary Mixtures

T. William Bentley\* and H. Carl Harris

Department of Chemistry, University College of Swansea, Singleton Park, Swansea SA2 8PP, Wales, United Kingdom

Received October 3, 1987

Rates and products for solvolyses of *p*-nitrobenzoyl chloride in aqueous binary mixtures with acetone, ethanol, and methanol are reported. Product selectivities,  $S = [\text{ester}][\text{water}]/[\text{acid}][\text{alcohol}]$ , increase in more aqueous media, in contrast to published data for solvolyses of benzoyl chloride. Logarithms of first-order rate constants are relatively insensitive to solvent ionizing power ( $Y$ ), showing marked dispersions for the three aqueous binary mixtures and unusual rate maxima for aqueous alcohol mixtures. Logarithms of calculated third-order rate constants for solvolyses in aqueous acetone (assuming these solvolyses are second order in water) show a linear Grunwald-Winstein plot ( $m = -0.18$ ), interpreted as the medium effect. Similarly, third-order rate constants for ethanol and methanol are calculated from solvolyses in the pure alcohols. Rate data in aqueous alcohols are explained quantitatively by an additional third-order pathway (first order in alcohol and in water), and product data are well explained if this process leads to ester. Thus mass law and medium effects of solvent molecules are separated, and it is shown that rate-product correlations of acceptable precision explain the unusual features of these solvolyses.

Classifications of reaction mechanisms in solution are usually based on the concept of molecularity: the number

of molecules necessarily undergoing covalency change during the rate-determining stage of the reaction.<sup>1a</sup> This