CHIMIA 40 (1986) Nr. 12 (Dezember)

REKEST: a Computer Program to REsearch for the KEy STep of a Synthesis – Application to Strategies for Building the Longifolene Skeleton**

René Barone* and Michel Chanon

Abstract: We describe the program REKEST (REsearch for the KEy STep) the aim of which is to research the key step of a synthesis by dissecting the target molecule according to formal schemes. These schemes represent generalized reactions which allow to delete and/or add one or several bonds in the target compound. By this method the user obtains rapidly a view of possible interesting intermediates. REKEST proposed new ideas for the synthesis of longifolene and recognized existing ones. REKEST is written in BASIC and runs on an Apple II microcomputer.

Introduction

There are presently a great number of computer assisted organic synthesis (CAOS) programs[1]. Most of these programs are developed according to the main principles defined by *Corey* and *Wipke* in their first paper[2]: they work backward from the target to the starting materials by construction of a synthetic tree. But this approach leads to a «combinatorial explosion» and the problem is to prune the synthetic tree. When a chemist plans a synthesis, he seldom works out such an extensive inventory. One may compare this problem to those associated with chess games, where computers analyze thousands of combinations whereas the great master concentrates rapidly on the best ones. As a consequence, the great masters regularly defeat the computer. So in these CAOS programs strategies and tactics have been elaborated in order to discard the bad solutions and to find the interesting ones faster[1].

Other approaches have been proposed to look for syntheses of a target compound. They differ from the classical approach because they solve only part of the problem, the chemist having then to complete the solutions. For example:

- *Wipke* proposed an approach which allows one to find the starting materials associated with a given target[3]; the chemist has then to find the best set of reactions which connects these reagents to the target.
- *Winter* proposed an approach to code synthetic pathways on the basis of strategies[4].
- SAS (Simulated Analytical Synthesis)[5,6] is based upon the analytical approach proposed by *Corey* et al.[7] and taken further by *Hamon* and *Young*[8]. The aim of this program is to cut bonds in the skeleton of the target molecule in order to display potential precursors.

The foregoing references reflect the diversity of the chemist's reasoning involved in the synthesis of a given target compound. The programmed simplifications result from the discovery of the key step in the considered synthesis.

Deslongchamp's article[9] on strategies, or the «chiron» approach developed by *Hanessian*[10] provide general illustrations of this notion of key step in synthesis planning. *Hendrickson*'s program involves initial dissection of the skeleton only, first to find all fully convergent modes of assembly, then the shortest sequences of constructions only to execute each[11].

The analytical approach allows one in some cases to find efficiently the key step of a synthesis. For example, SAS suggested the possibility of an intramolecular Diels-Alder reaction as the key step in the preparation of ellipticine, and a recent synthesis confirmed the soundness of this proposition[6,12]. However, SAS's approach is not actually comprehensive because it only breaks bonds: there is no formation of new bonds in the target compound; for exam-

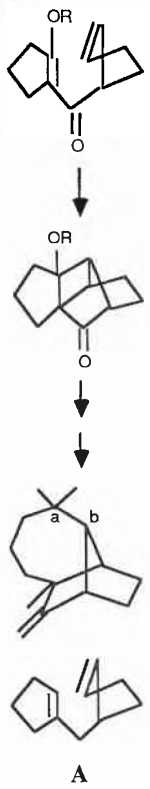
**Correspondence:* Dr. R. Barone
Laboratoire de Chimie Inorganique Moléculaire
U.A. CNRS 126
Faculté des Sciences de St. Jérôme
Rue Henri Poincaré, F–13397 Marseille cedex 13
(France)

***Acknowledgement:* The authors thank Prof. *J.B. Hendrickson,* Brandeis University, Waltham MA, for useful suggestions during the writing of this paper.

ple, rearrangement reactions cannot be envisaged by this program.

On the other hand, *Oppolzer's* synthesis of longifolene[13] shows that the concept of strategic bonds may be insufficient in the search for the key steps of a synthesis. In fact, *Oppolzer* et al. synthesize longifolene by an intramolecular DeMayo reaction, and one of the two bonds constructed (bond a–b) would not be considered as strategic according to the strategic bonds principle[13,14]. The key step is shown in Scheme 1.

Scheme 1



A

The finding of this key step, that is, of intermediate **A**, may be accomplished by a program working simply on the skeleton. Seeing the intermediate **A** printed should at once suggest to a good chemist the complete principle of this longifolene synthesis.

We describe in this paper the computer program REKEST (REsearch for the KEy STep) the aim of which is to discover efficiently the key step of a synthesis.

Description of REKEST

The main idea of the program REKEST is to provide a large number of formal solutions as propositions for the chemist. Instead of constructing a complete synthetic tree it points out some potential intermediates in the tree which may suggest a key step for the synthesis. It dissects the target with a small library of «reactions». These «reactions» are patterns which allow to delete and/or add one or several bonds in the target compound. The main «reactions» are depicted in Scheme 2 (A, B, C, D are any atoms; transformations 1 to 3 allow deletion of one or two bonds).

Scheme 2

1/ A-B	<-	A B
2/ A-B-C	<-	A B C
3/ A-B C-D	<-	A B C D
4/ A-B	<-	A-L + B
5/ A-B-C	<-	A + B=C
6/ A-B-C	<-	A-C + B
7/ A-B C-D	<-	A-D + B-C
8/ A B	<-	A-B

him to work either on a fragment of the target or on it as a whole. To avoid the generation of too many solutions, the chemist has the possibility of selecting the atoms which will participate.

Then the program reads one reaction and checks if it is applicable; if «yes», the transformations are performed and the corresponding solutions are presented to the chemist who may either save the intermediate, have a hard-copy of the screen, or continue; that is, it returns to the «is the reaction present?» instruction, and if it is

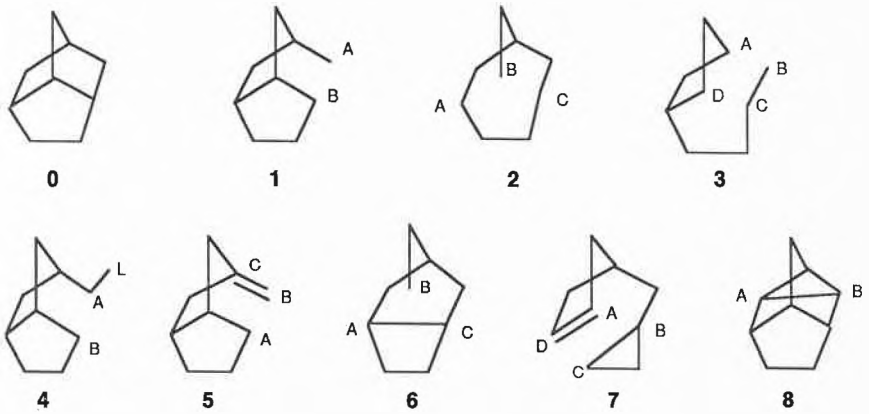


Fig. 1. Example of each transformation 1–8 (see Scheme 2) applied to target 0.

Scheme 3

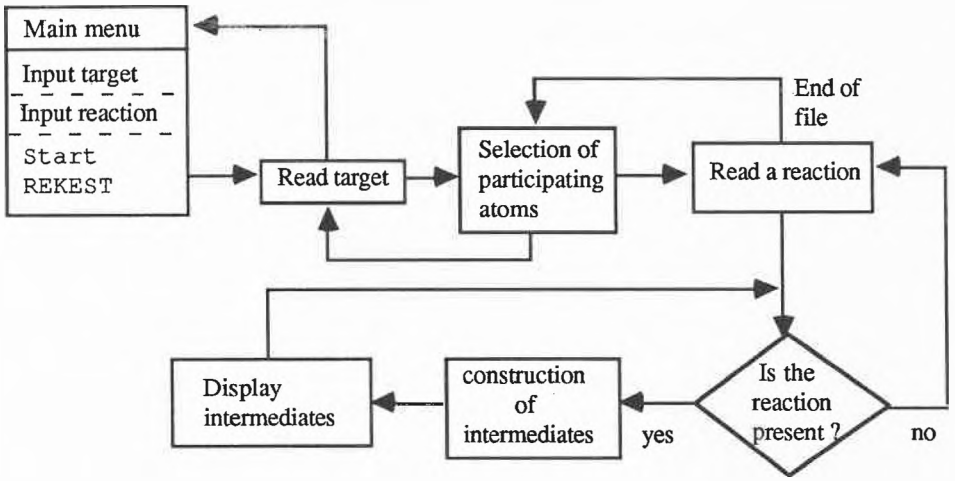


Fig. 1 shows an example of the application of each transformation on a given target (0).

The flow chart of the program is presented in Scheme 3. The main menu offers three options:

(1) Input and storage of the target compound: the user draws the target on the screen by means of the keyboard and the computer establishes the internal description (list of atoms and bonds) which is saved on disk.

(2) Input of «reactions».

(3) Execution of program REKEST.

REKEST begins by reading the target and draws it on the screen. An option is then provided to the chemist which allows not it reads another reaction and the sequence is repeated until the end of file.

When all the «reactions» have been tried, the program returns to the instruction «choice of atoms»: it is possible to select other atoms for researching other solutions. Of course it is also possible to select all the atoms in this step in order to be sure to have all the solutions from the given set of reactions, but in this case the number of solutions may be too large.

Results

We have tested REKEST with the longifolene skeleton and we present in Fig. 2 the solutions which seem interesting. The actual number of solutions depends on the number of atoms present in the selected

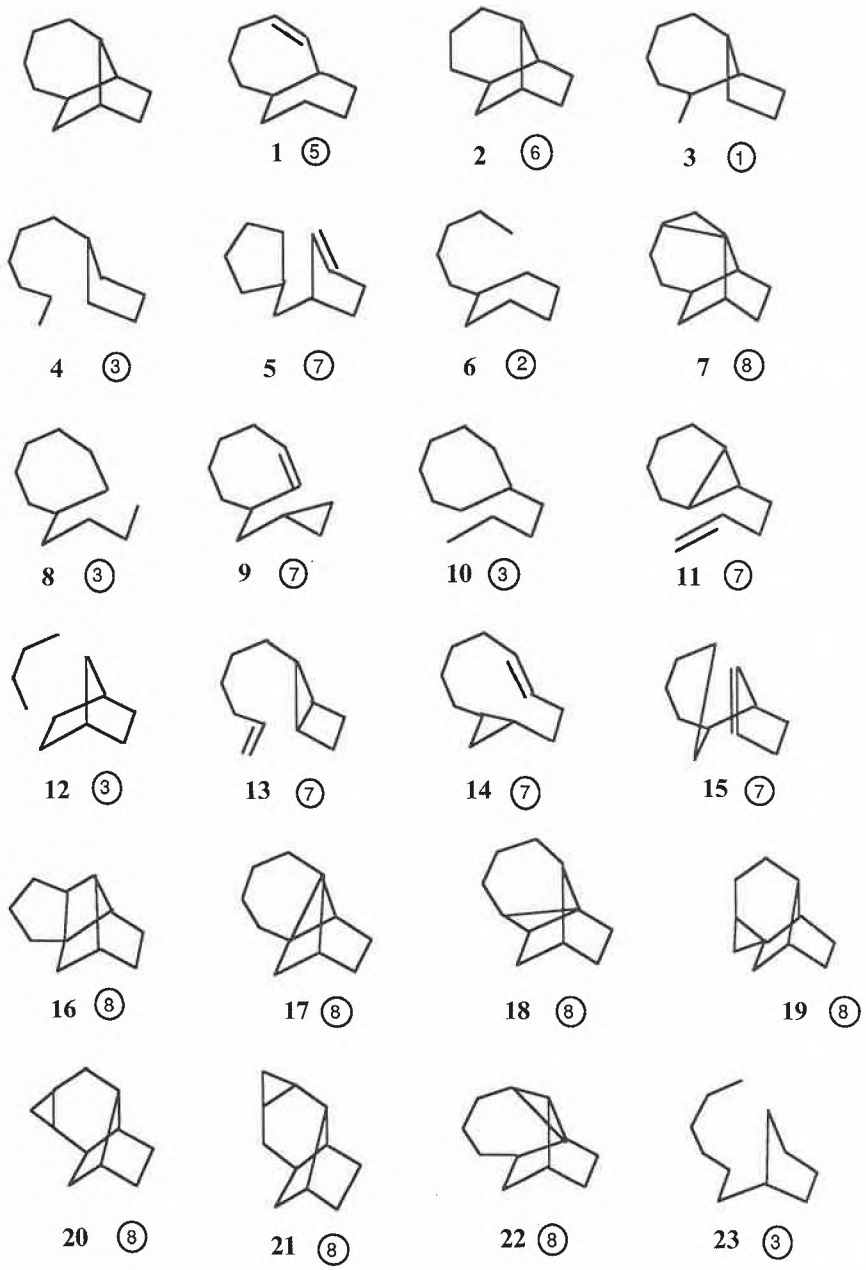
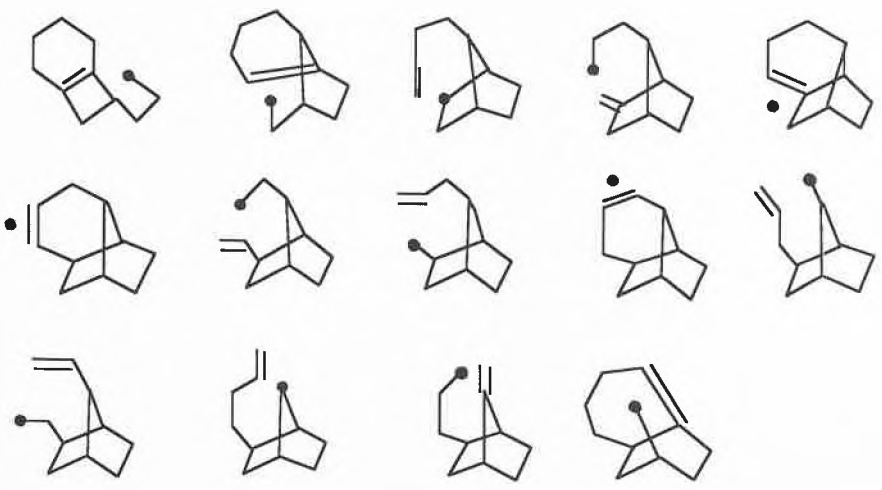


Fig. 2. Some results provided by REKEST for the longifolene's case. The circled numbers indicate the numbers of the «reactions» (Scheme 2) which generate the intermediates. With SAS[5,6] solutions **1, 2, 5, 7, 9, 11, 13, 14, 15, 17–22** *would have been missed.*

Scheme 4



fragment of the skeleton. For example, for the fragment formed by the seven-membered ring one obtains 112 solutions. For the whole skeleton there are 348 answers. The selection of solutions is made by the chemist, the aim of REKEST being to propose potential intermediates which could suggest the key step of the synthesis.

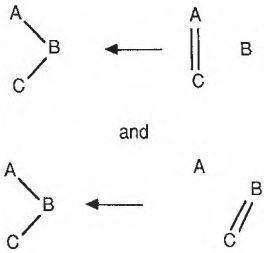
Intermediate **1** corresponds to the synthesis realized by *Corey* et al.[7], solution **2** to the one by *McMurry* et al.[15]. Structure **3** is the last step of *Johnson*'s approach[16]. Intermediate **4** suggests an intramolecular Diels-Alder reaction which was unsuccessfully tried by *Briegger*[17] but also the strategy retained by *Johnson* et al.[16]. Structures **5** and **16** suggest the synthesis involving the key intermediates imagined by *Oppolzer*[13]. Intermediate **6** suggests *Schultz*'s synthesis via an intramolecular diene-carbene cycloaddition[18]. Solution **7** shows the cyclopropane key intermediate of the synthesis by *Fallis* et al.[19].

Intermediates **8–11** suggest internal [3 + 2] cycloaddition like:

Syntheses of five-membered rings by this sort of reaction have been described[20,21]. However, here the solution is more challenging because a bridged structure has to be built using this principle.

The cyclopropane ring present in solutions **17** to **21** suggests carbene olefin precursors such as:

for the formation of **17**. Such a precursor may be found directly by adding the following «reactions» to the ones shown in Scheme 2, forming the selected basis for the foregoing target:



They represent the overall result of an addition of carbene to the double bond, followed by the opening of the cyclopropane ring. These reactions applied to the seven-membered ring give, among 21 products, the precursors of Scheme 4 (● stands for carbene).

These solutions correspond of course to rough propositions that the chemist must then examine critically.

We have purposely included several «silly» propositions to show that the common sense of the chemist demands discarding many of the intermediates proposed by REKEST. The other ones illustrate, however, what kind of aid REKEST can provide to chemists. Functionalized longifolenes may then be obtained through reactions described in *Jadhav*'s review[22].

Transformations given in Scheme 2 are not to be considered as comprehensive; the user can enter his own «reactions» into the file and thereby define his personal strategy.

Conclusion

Computer assisted organic synthesis is usually dealt with by programs which try to create the synthetic tree containing step by step all useful routes to the target molecule. We propose, with the program REKEST, a different, simpler approach to search only for the possible key steps of a synthesis by working on the skeleton of the target compound.

Some ot these key steps may then suggest new syntheses to the chemist.

REKEST is specially well adapted to treat bridged complex compounds with few chemical functions.

The program is written in BASIC and runs on an Apple II microcomputer; an IBM/PC version is under development. It can be made available from the authors.

Received: April 7, in revised form: October 28, 1986
[FC 61]

[1] R. Barone, M. Chanon, in G. Vernin, M. Chanon (Ed.): *Computer Aided Chemistry*, Ellis Horwood, Chichester (1986), p. 1.
[2] E.J. Corey, W.T. Wipke, *Science 166* (1969) 178.
[3] W.T. Wipke, D. Roggers, *J. Chem. Inf. Comput. Sci. 24* (1984) 71.
[4] J.H. Winter, *J. Chem. Inf. Comput. Sci. 25* (1985) 389.
[5] A. Boch, R. Barone, M. Chanon, J.M. Metzger, *Comput. Chem. 3* (1979) 83.
[6] R. Barone, M. Chanon, *Heterocycles 16* (1981) 1357.
[7] E.J. Corey, M. Ohno, R.B. Mitra, P.A. Vakatencherry, *J. Am. Chem. Soc. 86* (1964) 478.
[8] D.P.G. Hamon, R.N. Young, *Aust. J. Chem. 29* (1976) 145.
[9] P. Deslongchamp, *Bull. Soc. Chim. Fr. II (9–10)* (1984) 349.
[10] S. Hanessian, in J.E. Baldwin (Ed.): *The Total Synthesis of Natural Products: the 'Chiron' Approach*, Organic Chemistry Series, Pergamon, Oxford (1983).
[11] J.B. Hendrickson, D.L. Grier, A.G. Toczko, *J. Am. Chem. Soc. 107* (1985) 5228.
[12] E. Differding, L. Ghosez, *Tetrahedron Lett. 26* (1985) 1647.
[13] W. Oppolzer, T. Godel, *J. Am. Chem. Soc. 100* (1978) 2583.
[14] E.J. Corey, W.J. Howe, H.W. Orf, D.A. Pensak, G. Peterson, *J. Am. Chem. Soc. 97* (1975) 6116.
[15] J.E. McMurry, S.J. Isser, *J. Am. Chem. Soc. 94* (1972) 7132.
[16] R.A. Volkmann, G.C. Andrews, W.S. Johnson, *J. Am. Chem. Soc. 97* (1975) 4777.
[17] G.T. Briegger, *J. Am. Chem. Soc. 85* (1963) 3783.
[18] A.G. Schultz, S. Puig, *J. Org. Chem. 50* (1985) 916.
[19] S.K. Attah-Poku, K. Antczak, S.J. Alward, A.G. Fallis, *Can. J. Chem. 62* (1984) 1709, 1717.
[20] B.M. Trost, D.M.T. Chan, *J. Am. Chem. Soc. 104* (1982) 3733.
[21] B.M. Trost, *Angew. Chem. Int. Ed. Engl. 25* (1986) 1.
[22] P.K. Jadhav, U.R. Nayak, *J. Sci. Ind. Res. 39* (1980) 36.