Computer-Aided Organic Synthesis – SESAM: A Simple Program to Unravel “Hidden” Restructured Starting Materials Skeleta in Complex Targets

Goverdhan Mehta[a], René Barone*[b], and Michel Chanon[b]

School of Chemistry, University of Hyderabad[a],
Hyderabad 500134, India
E-mail: gmchem@uohyd.ernet.in

Laboratoire AM3, Fac. Sci. St. Jérôme, ESA CNRS 6009[b],
F-13397 Marseille cedex 20, France
E-mail: rene.barone@am3.u-3mrs.fr

Received February 3, 1998

Keywords: Computer-aided organic synthesis / Retrosynthetic strategy / Starting materials strategy / Retrosynthetic disconnections / Taxol

We describe the SESAM program whose aim is to help the chemist in exploring new synthetic strategies for complex structures by locating simple, non-obvious but effective starting materials. The program works at the skeletal level and attempts to map simply modified starting materials into the skeletal surface of the target. SESAM has been tested on taxol and steroid frameworks.

Introduction

One of the key elements in the planning of synthesis of complex molecules, particularly natural products, is the appropriate choice of the starting material (chiron, in case of chiral synthesis). If the synthetic endeavour involves construction of intricate carbocyclic skeleta, then it is imperative that the structural features of the starting material (SM) should be readily recognizable in the target structure. Identification of such relationship between the SM and the target structure has been routinely done by synthetic chemists through intuition and visual recognition of structural patterns. However, in 1969 Corey and Wipke introduced the use of computers to aid synthetic design[1] (CASD) and demonstrated the efficacy of retrosynthetic analyses in identifying the starting materials that bear some of the structural features of the target compounds. Thus, starting from the target, their LHASA program generates precursors which, in turn, become targets and new precursors are generated. The process is repeated until commercial or simple precursors are found and a synthetic tree is generated[2]. Several programs based on this approach have been developed[3]. Other strategies, like multiple constructions[4], search for the key step[5] and search for very appropriate starting materials have also been proposed. For example, Wipke was the first to introduce a program to search for SM inside the target from the databases of commercially available starting materials[6]. Similar approaches have been used again in SYNGEN[7], CHIRON[8], WODCA[9], and LHASA[10] programs.

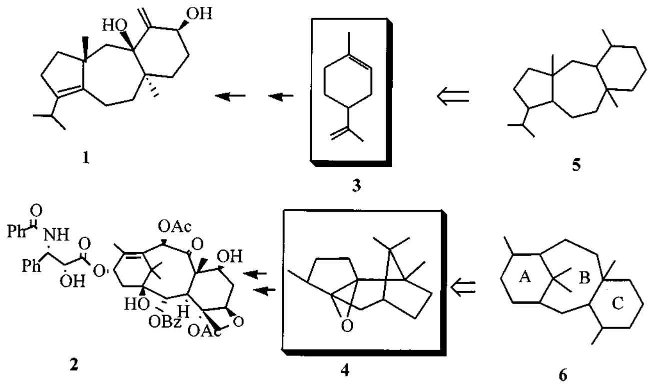
In SYNGEN[7] the target skeleton is cut into two pieces which are checked with a catalog of skeletons of SM. The process is repeated at two levels. When all the pieces have been found in the catalog, the program creates appropriate functions. The main feature of the CHIRON[8] program is the search for appropriate SM from an existing data base for synthesis. This search is done by a process of pattern recognition involving carbon skeleton, functional, and stereochemical convergences. Several options are available: general search in the database; searching by nature of carbon skeleton; searching by functional groups, searching for precursors to a specific substructure; searching by the number of stereogenic centers, by the number of carbon atoms; searching for chiral, achiral or racemic precursors. WODCA[9] is composed by a series of computer methods for synthesis design. One of the most important aspect of WODCA is the creation of relationships between a synthetic target or precursor and a starting material. WODCA provides several retrieval methods such as full structure search, fragment-oriented similarities measures, numerical and textual search on names and structures attributes. WODCA contains 48 different definitions of similarity which can be divided into two sets. The first one is based on perceiving common features in the substructures of the compounds to be compared with each other. The second one is based on reaction types, that is useful for the design of organic syntheses, since the program gives an idea of the reaction to be used to transform the SM into the target. One strategy developed in the LHASA program is the starting-material-oriented retrosynthetic analysis[10]: this strategy makes it possible to select potential SM and to perceive the structural changes between the target and the SM and then to find the reactions that can bring out these changes.

The presently available programs tend to locate *direct structural similarities* between the starting materials and the target. However, in many instances, particularly when terpenes are employed as chirons, *there are “hidden” structural*

© WILEY-VCH Verlag GmbH, D-69451 Weinheim, 1998 1434–193X/98/0707–1409 $ 17.50+.50/0

identities between the starting materials and targets. They are neither obvious to the eye nor indicated through computer analysis. Recent synthetic approaches to the diterpenes isoamijiol (**1**)[11] and taxol (**2**)[12] from (+)-limonene (**3**) and (−)-β-patchoulene oxide (**4**), respectively, are two examples that illustrate this point (Scheme 1). The presence of **3** and **4** in the skeleta **5** and **6** of **1** and **2**, respectively, is not quite apparent, yet **3** and **4** have served as excellent starting materials for **1** and **2**. This illustrates the importance of recognizing restructured starting materials into target structures for which no computer-based protocols are available. When such "hidden" identities are revealed, the starting materials can be suitably restructured through routine synthetic manipulations.

Scheme 1



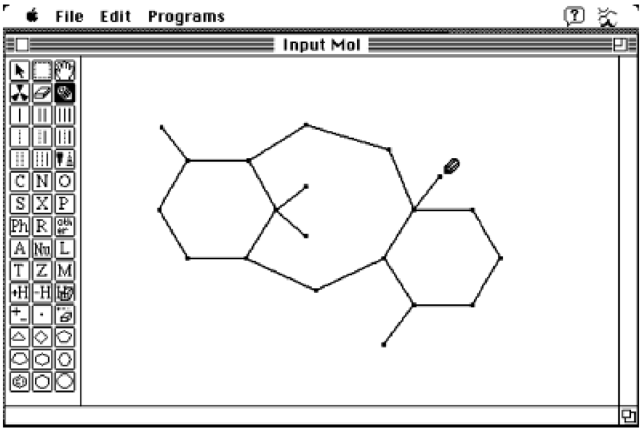
As part of our continuing efforts to provide the synthetic chemist with simple, user-friendly programs available on microcomputers[5][13], we report here the development of a new program that helps in locating restructured starting materials, particularly from terpenic chirons in the target structures which we reckon bestows enhanced versatility and manoeuvrability on the starting materials and provides new leads for complex synthesis. In this program we do not use databases of starting materials from commercial suppliers as in other programs[6][7][8][9][10], our aim was not to rewrite existing programs, but to offer a different and simple approach, available on microcomputer. This approach concentrates on the carbon skeleton only. The efficiency of such an approach has been shown in the field of CASD[5]. One must keep in mind, when using SESAM, that it does not deal with the stereochemical and the functional aspects of the target.

Program

This program, called SESAM (SEarch for StArting Material), has been developed on a Macintosh microcomputer. It works at a skeletal level: the target and the starting material are described by their skeletons. For example, if the target is taxol (**2**), it will be described by its skeleton **6** (Scheme 1). SESAM contains two modules: (i) one for drawing the skeleta of the target and the SM and (ii) the second to search for the SM skeleta in the target.

The drawing of structures is done by a user-friendly interface by means of mouse, menus, and icons. Figure 1 shows this part of the program. The user may save the structures on disk. Several options allow the user to manage the files.

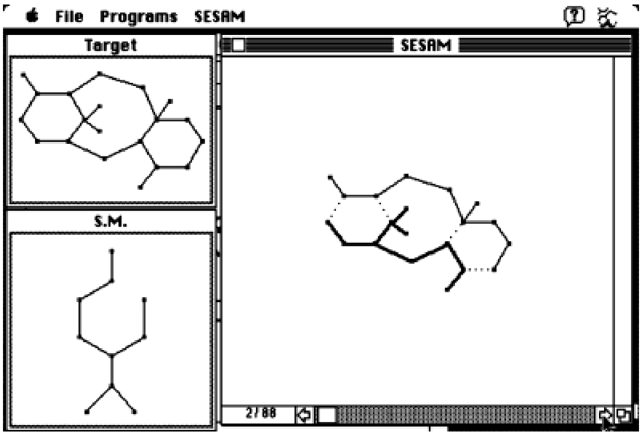
Figure 1. View of the screen during the input of the target skeleton



The target and the SM are described by connectivity tables and the program compares these tables by an atom/atom and bond/bond matching using a classical backtracking algorithm. The first bond in the SM is between atoms 1 and 2, and the bond order is 1 (all bonds are simple, since we work on the skeleton, but the subroutine is general and can find any substructure). Thus the program searches for an atom in the target identical to atom 1 of the SM. When it is found, it searches if this atom is bonded, by a simple bond, to an atom identical to atom 2, and so on. When a match is not found the program comes back to the last found atom and the search continues. The search is fast, on a Performa 6200 Macintosh, the search of an SM of 10 atoms in the taxol skeleton, and the sorting of 88 solutions is about 3 seconds and 0.26 seconds for the DOS version on a 300-MHz PC compatible.

For the "searching" module, the screen is divided into three windows: one for the target, one for the starting material and one for the results. When the user has selected the target and the SM, the program also offers the option of calculating the number of times the SM could be present in the target. Figure 2 shows the display on the screen during the search with the SESAM program.

Figure 2. View of the screen during the search with SESAM

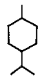
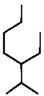
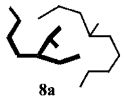
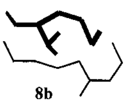
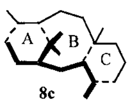
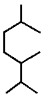
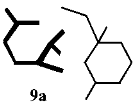
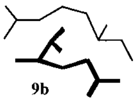
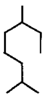
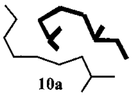
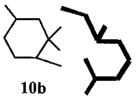
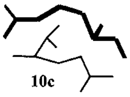
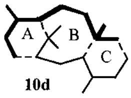
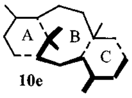
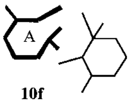
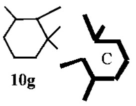
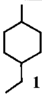
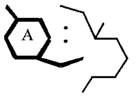


SESAM does not use commercially available databases of starting materials as indicated. The user is free to identify a particular starting material based on personal knowledge and preference, research interest, cost, ready availability, requirement of chirality and functionality, etc.

Results and Discussion

We illustrate the use and utility of SESAM through a few selected examples. For convenience, we have chosen examples from the terpene area, but the program is of general serviceability. Monoterpenes based on the *p*-menthane skeleton **7**, such as limonene, carvone, pulegone, menthone, etc. have often been employed as starting materials in synthesis, particularly of higher terpenes. While the choice of the SM in most of these cases has been intuitive, based on the recognition of structural patterns and requirements of functionality of the target, these mono-terpene precursors possess inherent ability to be restructured due to the functionalization present in them. Monoterpenes based on *p*-menthane (**7**) are known to undergo structural modifications through well-established, restructuring protocols available in the literature. Thus, skeleta **8−11** (with appropriate functionalization), among others, can be presumed to be readily accessible from **7**. As a result, a "basket" of starting materials based on skeleta of **7−11** is now available.

Table 1. Partial results for the search of some SM for the taxol framework

S.M.	Total number of solutions	Solutions involving some interesting dissections
 **7**	0	No solutions
 **8**	88	 **8a**  **8b**  **8c**
 **9**	41	 **9a**  **9b**
 **10**	89	 **10a**  **10b**  **10c**  **10d**  **10e**  **10f**  **10g**
 **11**	1	 **11a**

We now choose taxol (**2**), coded as **6** in the program, as the target and search the substructures (structural pattern) of SMs **7−11** present in it. Results are shown in Table 1. *With p-menthane* (**7**)*, the most commonly recognized skeleton among monoterpenes, there was no hit. However, when* **8** *(one bond removed in* **7***) was used as the SM, we were delighted to see that the program found it 88 times!* In order to handle such a large number of unexpected hits, the program was modified to display the solutions in increasing order of complexity, i.e., the number of components that will have to be united to construct the target molecule. The most interesting solutions are those which involve only two components (C_{10} + C_{10}); for **8**, the three solutions **8a−c** are displayed in Table 1. Table 2 summarises all the multicomponents solutions elucidated by the program[14].

Table 2. Number of solutions found for **8−11**, based on number of components involved; for example, for SM **8**, there are 3 solutions involving 2 components, 14 solutions involving 3 components, 21 solutions involving 4 components, etc.

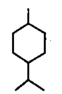
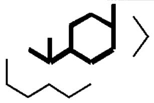
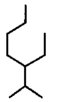
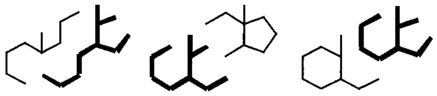
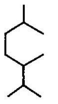
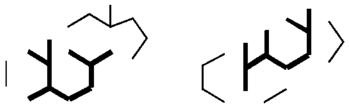
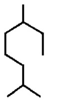
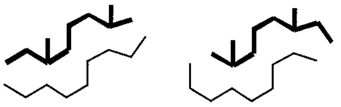
	Number of component							
SM	2	3	4	5	6	7	8	9
8	3	14	21	25	15	9	1	0
9	2	4	8	9	11	5	3	0
10	7	5	27	18	18	8	9	3
11	0	0	1	0	0	0	0	0

Similarly, substructure **9** is present 41 times in **6** with **9a−b** representing the two-component solutions. The substructure **10,** which incidentally is naturally occuring, is present 89 times with as many as seven (**10a−g**) two-component solutions. The C_9 substructure **11** is hit only once (**11a**) on the taxane framework **6** and the proposed solution involves a 4-component union. Interestingly, an approach along this line to the taxane A ring has recently been reported[15].

The results in Tables 1 and 2, provide a quite dramatic illustration of our program that is designed to locate the restructured starting materials on the target molecule. While the skeleton based on *p*-menthane (**7**) is totally unserviceable for taxanes (0 hit), its simple restructured forms **8−11** (one C−C bond cleavage in **7**) provide a multitude of solutions (219 hits). Some of these are insightful and provide new synthetic leads. For example, in **8c** the C_{10} unit derived from **8** is well poised towards a double Diels-Alder reaction (see dotted lines) for the framework construction with the diene and the dienophile moieties imbedded into it. Similar possibilities exist with **10d** and **10e** representing the fragment **10**. Noteworthily, in **10d** and **10e** the taxane framework is dissected into two identical C_{10} units, which could be combined through an inter- and intramolecular [4+2] cycloaddition cascade[16]. In **10f** and **10g**, the pattern of presence of substructure **10** shows that both rings A and C of taxanes can be constructed from the same C_{10} unit. Indeed, attempts to prepare ring A of taxane based on substructure **10** have been reported[17].

To further test the usefulness of SESAM, we applied it to the parent tetracyclic steroid skeleton. Monoterpenes have not been generally employed for the steroid synthesis and we thought of exploring the common, abundantly available skeleton of *p*-menthane (**7**) towards this end. When the skeleton of **7** was traced by the program in the steroid framework, there were seven hits, none of them involving, however, a simple two-component coupling (Table 3). When restructured precursors **8−10** were employed there was a major leap in the possible solutions, including some with two components. These are shown in Table 3. Once again the program found that modification/restructuring of the starting material (in this case **7**) had opened many new and non-obvious options.

Table 3. Partial results for the search of some SM for the steroid framework

S.M.	Total number of solutions	Solutions involving some interesting dissections
 **7**	7	
 **8**	65	
 **9**	33	
 **10**	87	

The program SESAM has been tested on a range of other target molecules employing isoprene, α-pinene, δ^3-carene, camphor etc., as the starting materials. The restructured forms of these terpene synthons provided an enormously large number of interesting solutions, thereby substantiating the novelty of our approach and the general utility of the program[18].

Conclusion

SESAM is a new program to help the chemist *in identifying "hidden" precursors for the design of complex structures.* By using simple, restructured forms (usually through C−C bond cleavage) of commonly used building blocks to map the skeletal surface of the target, many new synthetic leads are revealed. By employing a user-friendly interface at the skeletal level the search is fast and the chemist can conveniently and rapidly scan the various synthetic possibilities on the screen and select the most appropriate starting material.

[1] E. J. Corey, W. T. Wipke, *Science* **1969**, *166*, 178−192.

[2] E. J. Corey, W. T. Wipke, R. D. Cramer III, W. J. Howe, *J. Am. Chem. Soc.* **1972**, *94*, 421−430.

[3] [3a] M. Chanon, R. Barone, *Computers Aids in Chemistry* (Eds.: G. Vernin, M. Chanon), Ellis Horwood, Chichester, **1986**, 19−102. − [3b] G. Loftus, *Chemical Information Systems. Beyond the Structure Diagram*, Ellis Horwood, New York, **1991**, 222−262. − [3c] M. A. Ott, J. H. Noordik, *Recl.Trav. Chim. Pays-Bas* **1992**, *111*, 239−246, and the entire issue 6 of this volume.

[4] J. B. Hendrickson, P. Huang, *J. Chem. Inf. Comput. Sci.* **1989**, *29*, 145−151.

[5] [5a] G. Mehta, P. Azario, R. Barone, M. Chanon, *Tetrahedron* **1989**, *45*, 1985−1994. − [5b] G. Mehta, R. Barone, P. Azario, F. Barberis, M. Arbelot, M. Chanon, *Tetrahedron* **1992**, *48*, 8953−8962. − [5c] F. Barberis, R. Barone, M. Chanon, *Tetrahedron* **1996**, *52*, 14625−14630.

[6] W. T. Wipke, D. Rogers, *J. Chem. Inf. Comput. Sci.* **1984**, *24*, 71−80.

[7] [7a] J. B. Hendrickson, D. L. Grier, A. G. Toczko, *J. Am. Chem. Soc.* **1985**, *107*, 5228−5238. − [7b] J. B. Hendrickson, *ACS Symp. Ser.* **1996**, *626*, 214−231.

[8] S. Hanessian, *Pure Appl. Chem.* **1993**, *65*, 1189−1204.

[9] [9a] J. Gasteiger, W.D. Ihlenfeldt, P. Röse, *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 270−290. − [9b] M. Wagener, J. Gasteiger, *Angew. Chem. Int. Ed. Engl.* **1994**, *33*, 1189−1192; *Angew. Chem.* **1994**, *106*, 1245−1248. − [9c] W. D. Ihlenfeldt, J. Gasteiger, *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 2613−2633; *Angew. Chem.* **1995**, *107*, 2807−2829.

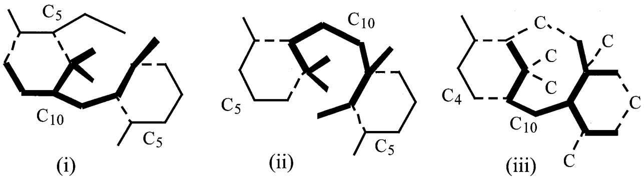
[10] A. P. Johnson, C. Marshall, P. N. Judson, *Recl. Trav. Chim. Pays-Bas* **1992**, *111*, 310−316.

[11] G. Mehta, N. Krishnamurthy, S. R. Karra, *J. Am. Chem. Soc.* **1991**, *113*, 5765−5775.

[12] [12a] R. A. Holton, R. R. Juo, H. B. Kim, A. D. Williams, S. Harusawa, R. E. Lowenthal S. Yogai, *J. Am. Chem. Soc.* **1988**, *110*, 6558−6560. − [12b] R. A. Holton, H. B. Kim, C. Somoza, F. Liang, R. J. Biediger, P. D. Boatman, M. Shindo, C. C. Smith, S. Kim, H. Nadizadeh, Y. Suzuki, C. Tao, C. P. Vu, S. Tang, P. Zhang, K. K. Murthi, L. N. Gentile, J. H. Liu, *J. Am. Chem. Soc.* **1994**, *116*, 1599−1560

[13] P. Azario, R. Barone, M. Chanon, *J. Org. Chem.* **1988**, *53*, 720−724.

[14] There are some interesting multicomponent solutions like **8→i** (C_{10} + C_5 + C_5) and **10→ii** (C_{10} + C_5 + C_5) which involve union of the C_{10} unit with two identical C_5 units. The possibilities of employing the [4+2] cycloaddition for the framework construction are indicated (dotted lines in i and ii). Less interesting solutions like (**8→iii**) involving as many as eight components have not been shown, but all such solutions are tabulated in Table 2.



[15] A. Srikrishna, T. J. Reddy, P. P. Kumar, *J. Chem. Soc., Chem. Commun.* **1996**, 1369−1370.

[16] [16a] T. Doi, J. Robertson, G. Stork, A. Yamashita, *Tetrahedron Lett.* **1994**, *35*, 1481−1484. − [16b] L. Alcaraz, J. J. Harnett, C. Mioskowski, T. Le Gall, D. S. Shin, J. R. Falck, *J. Org. Chem.* **1995**, *60*, 7209−7214.

[17] Y. F. Lu, A. G. Fallis, *Can. J. Chem.* **1995**, *73*, 2239−2252 and references cited therein.

[18] We will make available this simple computer aid to the interested persons among the organic chemistry community to enable them to explore many new synthetic leads. A DOS version for PC compatibles is also available (contact R. B).

[98041]