Tetrahedron Vol. 48, No. 18, pp. 3789-3804, 1992
Printed in Great Britain

0040-4020/92 $5.00+.00

COMPASS Program - An Original Semi-Empirical Approach To Computer-Assisted Synthesis

Ekaterina V. Gordeeva, Dmitri E. Lushnikov, Nikolai S. Zefirov*

N.D.Zelinsky Institute of Organic Chemistry,
Leninsky Prospect,47,117913,Moscow,USSR

(*Received in UK* 9 *March* 1992)

Key Words: computer-assisted synthesis; bond dissections; skeleton transformations; knowledge base; IBM PC

Abstract: *The fundamental principles of the original "DIssection-Approximation" approach (DIA-approach) to the formal description of organic reactions are presented. It is shown that organic reactions can be considered as dissections of chemical bonds followed by the appropriate "approximations" of the reaction centers around the dissected bonds. A new original COMPASS program for computer-assisted organic synthesis (CAOS) has been developed for IBM PC on the basis of DIA-approach. The COMPASS program is shown to be a reasonable compromise between two traditional (empirical and nonempirical) directions in CAOS. We have considered the approach which has been developed as an individual semi-empirical direction in CAOS. The main feature of this semi-empirical approach is shown to be the combination of pure combinatorial methods with a small knowledge base that includes empirical rules in their most generalized form. The powerful ability of the COMPASS program to reproduce the key skeleton transformations in the total syntheses of natural compounds is demonstrated for the structures quadrone and sirenin.*

INTRODUCTION

The history of CAOS development over the last two decades demonstrates that hereto CAOS is divided into two distinctly separated areas, namely, the empirical and non-empirical programs. Careful study of the fundamental principles of CAOS programs[1-4] clearly shows that the most significant point in the program development is the process of empirical knowledge formalization. The informal nature of organic chemistry makes it difficult to use the computer to aid in reproducing the logical, empirical, and intuitive considerations which guide the organic chemist in planning synthesis "by hand". Two principally different approaches in mathematical chemistry have been developed for the systematic formalization of knowledge about organic reactions. Thus, two directions in CAOS were developed namely the empirical and nonempirical approaches.[1-4]

Empirical CAOS historically arose earlier. The first program in this field (OCSS program) was described in 1969.[5] All the programs in empirical CAOS are based on a rather large kind of reaction library, and transformations of the target structures are implemented only in accordance with known organic reactions, which are preliminary coded and collated in the reaction library.[5]

The advantage of the empirical approach is that the program predicts synthetic methods that are usually reliable, and the selection criteria are contained directly in the description of each specific reaction.[6,7] On the other hand, such programs are restricted by a specific (and often quite large) reaction library and are not likely to suggest principally new synthetic pathways.[8] Moreover, the description of reactions in empirical CAOS programs is usually very detailed and, as a result, these programs appear to be oriented mainly to the investigations of synthetic methods rather than synthetic strategies.

Nonempirical CAOS arose later. The nonempirical approach is primarily used to generate transformations of target structures by means of logical-combinatorial procedures, using a small set of logical instructions or combinatorial algorithms. As a rule, nonempirical CAOS does not use empirical information about definite chemical reactions.

Programs of nonempirical direction are able to generate many new and original solutions. However, it is necessary to include in the program rigorous selection criteria in order to avoid a "combinatorial explosion" or to eliminate unreal and/or uninteresting results which can be generated during program operations.

Some of the strong and rather effective selection criteria were successfully introduced in nonempirical CAOS programs, and some interesting results were generated after application of the principle of minimum chemical distance,[9] enthalpy evaluation,[10] strategy of "ideal synthesis",[11] and restrictions based on the type of cyclic bond redistribution.[12,13] Unfortunately, the application of rigorous selection criteria to nonempirical CAOS programs, causes these programs to lose much of their nonempirical essence, and the set of generated solutions often appears to be incomplete from the view point of an experienced chemist. Moreover, such restricted programs are effective only for the simulation of rather narrow classes of chemical processes.

Another disadvantage of the nonempirical CAOS approach is that programs of this type are, as a rule, able to generate transformations only in closed systems in which the material balance is constantly maintained.[12-15] Hence, the nonempirical CAOS programs are usually not suitable for planning multistage organic syntheses. Indeed the user must have a prior knowledge of, and insert in a clear form to the program, not only the target structure but also a complete set of possible synthesis by-products. For example, if small molecules (water, carbon dioxide, hydrogen chloride, etc.) are eliminated in the synthetic process, the nonempirical CAOS program requires to know the exact numbers of these molecules in order to obtain one molecule of the target compound.

Previously we created the nonempirical FLAMINGOES program.[12] It has proved to be a good tool for the investigation of cage structure syntheses as well as various rearrangements and mechanisms of organic reactions.[13] The successful results of the FLAMINGOES program even surpassed our expectations. Further, the very promising application of this program to the investigation of reaction mechanisms initiated the development of the special combinatorial ICAR program for simulation of multistage carbocationic rearrangements.[16]

However, it was our goal to create a useful tool for the routine work of the organic chemist as FLAMINGOES is oriented "to the future", to search for principally new reactions and promising rearrangement processes.

The general tendency of CAOS development, as careful analysis of the literature shows, is the rapprochement of the empirical and the nonempirical directions of CAOS. As mentioned above, the saturation of nonempirical programs with strict selection criteria is sometimes effective. It should be particularly noted that these selection criteria usually have just an empirical background. Let us consider some examples of empirical rules, used for the selection of generated transformations. For instance, the EROS program selects the transformations by value of enthalpy, which is calculated empirically by means of an additive scheme.[10,17] The AHMOS program utilizes empirically defined "tendency of reactivity",[18] while TOSCA considers correspondence between heuristically determined consonant and dissonant strategies.[19] The

CAMEO program takes into account empirical values of nucleophilic and electrophilic properties of reaction centers.[20,21] The FLAMINGOES program generates transformations only for empirically selected reaction centers and applies the restrictions on the type of cyclic bond redistribution.[12,13]

On the other hand, a need for nonempirical (formal and exhaustive) consideration of possible solutions is now usual in the programs of traditional empirical approach. One can illustrate as an example the series of papers by Barone and Chanone, devoted to SOS and REKEST programs.[22-24] In these works the main point stressed is the search for selected key steps in the planning of an organic synthesis. The search is based on exhaustive enumeration of precursors, which are generated after application of a small set of effective skeleton transformations (e.g. retro-analogies for the De Mayo, oxy-Cope reactions, etc.).

However, the tendency to join the empirical and nonempirical approaches is distinctly traced in Hendrickson's logic-based program for synthesis design (SYNGEN).[25] This program involves many common features of the "classical" nonempirical approach, such as the non-interactive mode of operation and rigorously defined sequence for the construction of a "synthesis tree". The precursor generation is ordered in accordance with combinatorial enumeration of the so-called fully convergent bondsets.

Concerning the nonempirical essence of SYNGEN it should be noted that the generation of transformations (so-called half-reactions) is made on the basis of a restricted set of logical instructions describing all possible alternations on the immediate environment of the chosen skeleton atoms.

On the other hand, the latest versions of the SYNGEN program[25-26] take into account various restrictions which evaluate the applicability of individual half-reactions. Any half-reaction can be applied that is independent of the displacement of certain functional groups (FG). The possible electronic effects, induced by the corresponding FG, are also taken into consideration in Hendrickson's approach. Thus, there are some typical empirical features in the SYNGEN program.

Nevertheless the detailed analysis of a recent situation in the CAOS convinced us that there is no ideal program for the strategic planning of a multistep organic synthesis. Although, it is noteworthy that the computer-assisted planning of a complicated synthesis was declared as a main goal of CAOS project in the very beginning of the "computer age" in organic chemistry.[5]

We believe that the optimal CAOS programs should, in fact, possess the following necessary features (the requirements are ordered in decreasing importance) :

(1) The program should be able to generate complicated multistep synthetic pathways to a given target structure;

(2) The tactical details of an individual synthesis stage implementation should be omitted in order to avoid the "combinatorial explosion" while enumerating all possible FG interchanging transformations (FGI-transformations);

(3) The basic empirical information about organic structures and reactions should be taken into account in order to exclude the generation of unreal or naive solutions; the possibility of reproducing important routine synthetic methods should be maintained;

(4) The abstract representation of a chemical reaction should be applied in order to generate some new and original solutions while keeping in accordance with requirement (2);

(5) The program should consist of combinatorial algorithms and modules in order to keep the impartiality in searching for new results, as the nonempirical CAOS programs do.

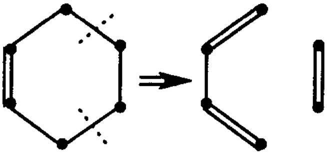
Thus, the challenge of generating a principally new version of the CAOS program arises which should be initially based on compilation of nonempirical exhaustive combinatorial enumeration of solutions and on the open system of simple empirical rules. This rule system should describe in the most generalized form all possible transformations which could be applied for the generation of precursors of the target structure. In our opinion it is now feasible to elaborate on the general principles of the new approach in CAOS based, for example, on the requirements (1)-(5). We also suggest the use of a special term to mark this combinatorial approach as the Semi-Empirical Computer-Assisted Organic Synthesis.

In this paper we present the general principles and typical results of our COMPASS program which implements, as fully as possible, the semi-empirical ideology of CAOS.

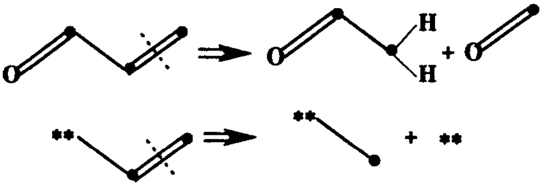
DESCRIPTION OF REACTIONS IN COMPASS PROGRAM (DIA-APPROACH)

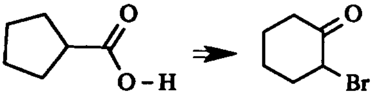
In order to represent the chemical retro-reactions (transformations) in the COMPASS program, we have elaborated, in accordance with requirements (3) and (4), a special approach to describe structural changes in terms of "dissections" and "approximations" and we abbreviated this method the DIA-approach for brevity.

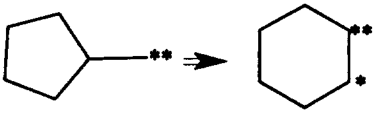
In the DIA-approach the individual chemical transformation is described as a dissection of one or two bonds in the chemical structure with subsequent "approximation" (saturation) of free valences, which are formed formally during bond cleavage. All possible variants of the approximations can be classified generally as transformations of the formally generated free valences to σ- or π-bonds (in the DIA-approach: σ- or π-approximation, respectively). The well-known Diels-Alder reaction (in the retro-direction) is a good example for a π-approximation:



In turn, the σ-approximations can be subdivided into two classes: F-σ-approximations and R-σ-approximations. F-σ-approximations (Functional σ-approximations) are generated as a result of FG addition,while R-σ-approximations (Rearrangement σ-approximations) are interpreted as σ-bond formation in a rearrangement process.





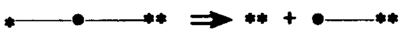
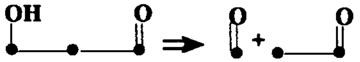
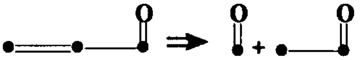
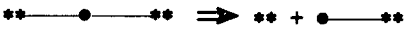
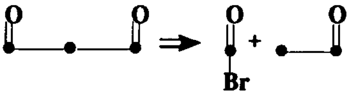
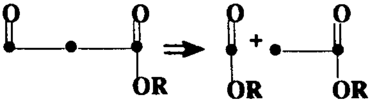
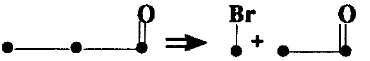
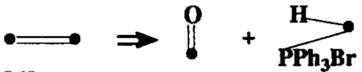
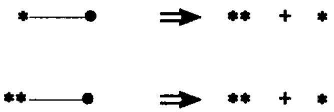
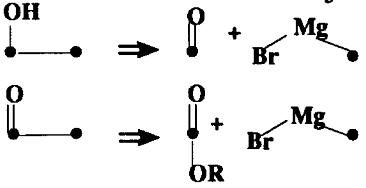
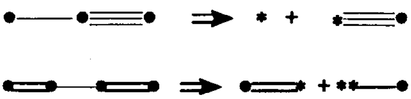
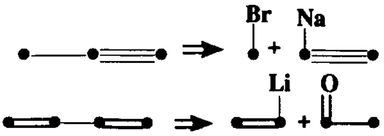
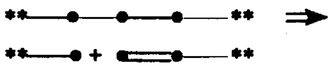
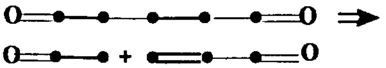
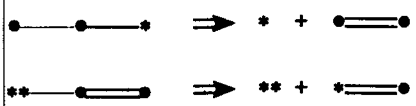
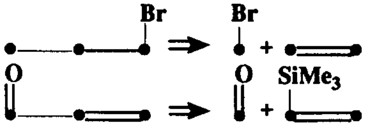


The DIA-approach is oriented mainly to the description of skeleton transformations. The FG are of minor importance and are described in an abstract manner in terms of Activated Centers. The carbon atom connected with one heteroatom by an ordinary bond is considered to be a Monoactivated Center (marked by one asterisk). The carbon atom connected to more than one heteroatom or connected to one heteroatom by a multiple bond is considered to be a Diactivated Center (marked by two asterisks). The hierarchy of dissection and approximations, introduced in the DIA-approach, allows attention to be concentrated on the search for the strategy of a multistep synthesis, the key stage of formation of the target molecule skeleton.

In accordance with the logic of an organic chemist who is using a strategic plan for an organic synthesis, we believe FGI- transformations to be of minor importance. However, it is noteworthy that the DIA-approach does not exclude entirely FGI transformations from consideration. If necessary, transformations of this type also can be considered (with a minor rating) as a replacement for activated centers around the dissected bond.

In the present version of the COMPASS program approximations are given as a set of simple generalized empirical rules. Tables 1 and 2 show the rules which are applied by the COMPASS program for

Table 1. Rules for F-σ-approximations included in the COMPASS program

DIA-rule	Synthetic analog	Forward reaction
		aldol condensation
		crotonic condensation
		acylation
		ester condensation
		alkylation
		Wittig reaction
		Grignard reaction
		acetylenides
		Michael addition
		addition to double bond

F-σ-approximations after dissection of carbon-carbon and carbon-heteroatom bonds, respectively. The formal description of subsequent dissection and approximation (so called DIA-rule) is presented in the left column. The central column contains the detailed DIA-rule with examples of specific FG. The FG, chosen for illustration in Tables 1 and 2, are usually applied in practical chemistry for a given type of skeleton transformation. The traditional name of the corresponding forward organic reaction is given in the right column. F-σ-approximations make up the most numerous class of approximations in the COMPASS program.

π-Approximations (Table 3) are illustrated by retro-reactions of cycloaddition, carbene addition and electrocyclic ring formation. For the application of π-approximations, the cyclic fragment should be found in the target molecule. If this requirement is satisfied, the COMPASS program immediately generates the

Table 2. F-σ-approximations for carbon-heteroatom bonds in the COMPASS program

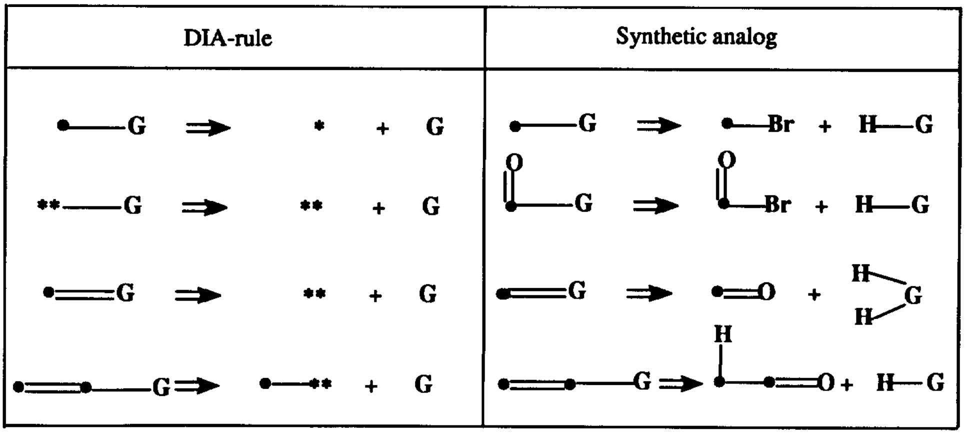
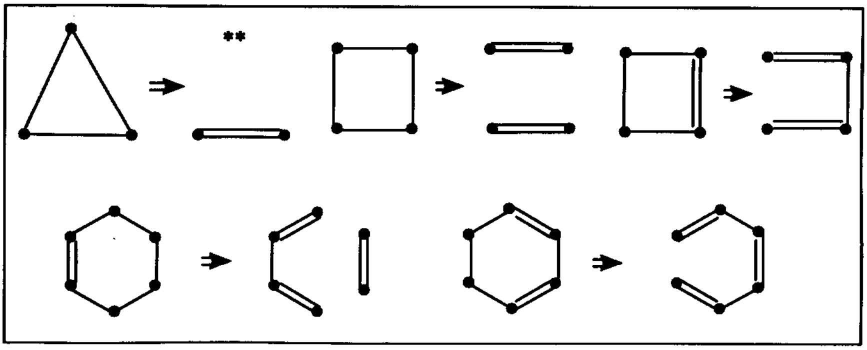


Table 3. Rules for π-approximations

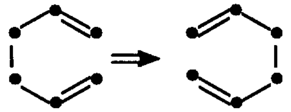
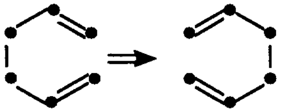
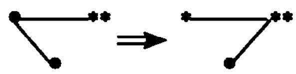
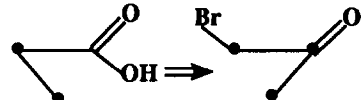
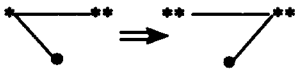
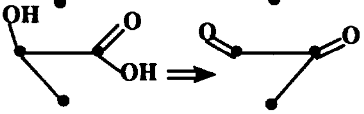
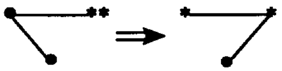
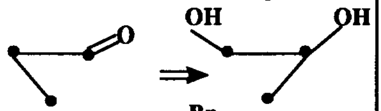
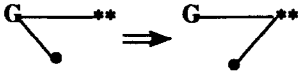
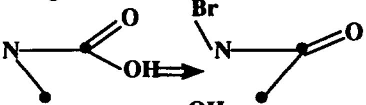
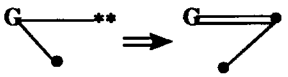
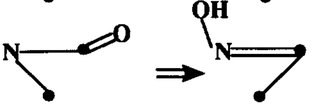
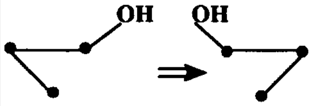
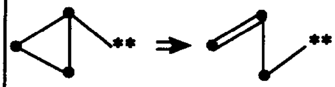
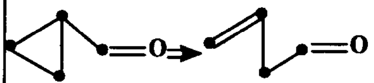
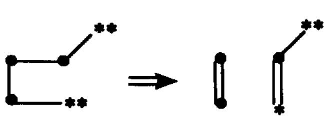
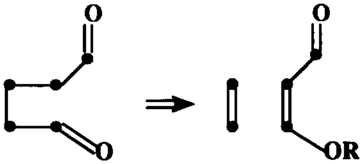


dissections and the corresponding π-approximations of one or two bonds belonging to 3-6 membered rings. It should be noted that in π-approximations the description of the favorable displacement of activated centers is not defined which is in contrast to the description of F-σ-approximation. In other words, we believe π-approximations to be such powerful and profitable retro-synthetic steps, that they should be applied for precursor generation independent of the specific FG disposition in accordance with requirement (2) (see above).

Very interesting retro-synthetic pathways can be found on application of R-σ-approximations to the dissection of the target structure[5]. Table 4 shows the DIA-rules for R-σ-approximations used presently in the COMPASS program. It should be mentioned that a generalized description of dissection and approximation by the DIA-rule allows the numerical set of transformations to be collected under the same rule with the common type of skeletal changes. For instance, the DIA-rule for the [3,3]-sigmatropic rearrangement is applicable not only for the "classical" variant of the Cope rearrangement, but also for the oxy-Cope reaction and for the Claisen rearrangement.

Most of the R-σ-approximations in Table 4 represent the various types of 1,2-nucleophilic migrations. In order to distinguish these rearrangements as accurately as possible, and to make it easier for a chemist to interpret the results, a more detailed description of the activated center is introduced. The diactivated centers are additionally classified in accordance with the total number of valences which connect a given carbon

Table 4. Rules for R-σ-approximations in the COMPASS program

DIA-rule	Synthetic analog	Forward rearrangement
		[3,3]-sigmatropic
		Favorski
		benzylic
		pinacolinic
		Hoffman
		Beckman
		Wagner-Meerwein
		oxa-di-π-methane
		De Mayo

atom with heteroatoms. Thus, the description of R-σ-approximation subdivides the diactivated centers to "real" diactivated centers (e.g. carbonyl group) and triactivated centers (e.g. carboxyl group).

Unfortunately, the mutual displacement of the activated centers and multiple bonds are not always favorable for the direct application of a given DIA-rule. Hence, it is often necessary to forestall the application of an approximation by preliminary modifying the activated centers around the dissected bond. A similar situation is often encountered in laboratory practice, where it is necessary to carry out the sequence of FGI transformations before the implementation of an effective constructive[27] reaction.

The rules for Activated Center Modification (M-rules) used in the present version of the COMPASS program are collected in Table 5. All of these transformations correspond to the simplest changes in a chemical structure which proceed without skeletal modifications. In Table 5 the M-rules are interpreted as a keto-enol tautomeric transformation, the addition/elimination of FG to/from a multiple bond, the reduction/oxidation of an individual carbon atom, etc. It should be emphasized that in the M-rules, as well as

Table 5. M-rules in the COMPASS program

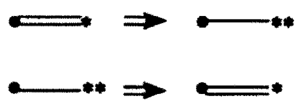
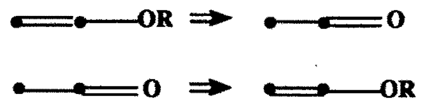
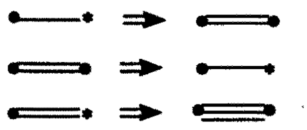
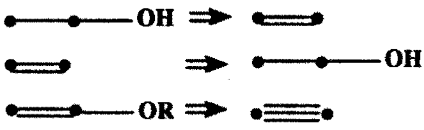
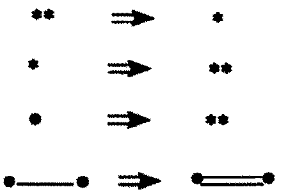
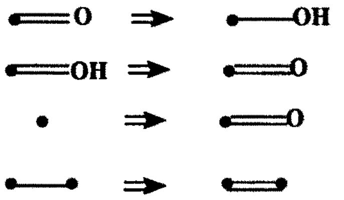
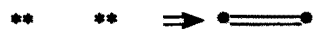
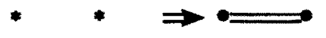
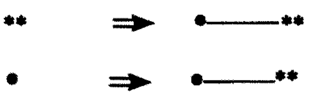
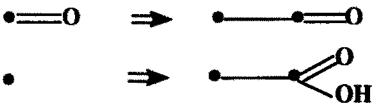
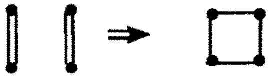
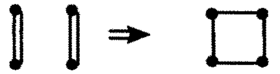
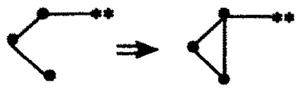
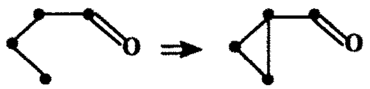
M-rule	Synthetic analog	Forward reaction
		Keto-enol tautomerism
		Addition-elimination
		Oxidation-reduction
		Epoxidation

Table 6. Skeleton complication rules in the COMPASS program

C-rule	Synthetic analog	Forward reaction
		Ozonolysis
		Reductive cleavage
		Decarboxylation
		Cyclo-fragmentation
		Reductive cleavage

in the R-σ-approximations, the di- and triactivated centers should be distinguished.

It is well-known that an organic synthesis cannot be always represented strictly as a subsequental complication of initial structures in the creation of the target compound. There are many examples of elegant syntheses (see, for instance,[28-29]), in which some key stages are just a simplification of the precursor skeleton. In this situation the target compound has a simpler structure than its precursor. To treat such cases in retro-synthetic planning the preliminary skeletal complication should be carried out before applying a certain approximation. The special Complication rules (C-rules) are presently used in the COMPASS program (see Table 6), and these transformations are applied before the DIA-rules (like the M-rules). The generalized C-rules (used in the COMPASS program) are illustrated in Table 6 by using, as examples, retro-ozonolysis, oxidative double bond retro-cleavage, decarboxylation, decarbonylation, [2+2]-cyclofragmentation and cleavage of the 3-membered ring after an oxa-di-π-methane rearrangement.

COMPUTER-ASSISTED REPRODUCTION OF STRATEGIC STEPS IN TOTAL SYNTHESES OF NATURAL COMPOUNDS

The structure of quadrone (a well-known natural sesquiterpene compound) is a convenient example for testing the CAOS program. Indeed, more than 10 total quadrone syntheses have been published since 1981. Therefore, the reality of retrosynthetic paths predicted by the computer can be correctly evaluated by the comparison of "synthesis trees" generated from quadrone with the synthetic routes published.[27,30-41]

Unfortunately, the limited space of a journal publication does not allow the presentation of all the total syntheses of quadrone reproduced by the COMPASS program. Only five typical examples are demonstrated and discussed.

It should be mentioned that the COMPASS program reproduces primarily the key stages of the target skeleton formation. The numerical FGI transformations are often omitted. It also should be taken into account that all FG in the framework of the DIA-approach are described generally in terms of activated centers. Specific atom symbols are introduced in the program output to allow visualization of results, presenting output in the most readable form.

To reproduce the retrosynthetic pathways most closely related to the published ones the COMPASS program was run in semi-automatic mode: if the structure involved in the experimental synthetic route was encountered among the computer-generated precursors, it was manually selected as a synthetic subgoal for a further retrosynthetic search. Other structures generated on the same level of the "synthesis tree" were discarded at this point. However, between the subgoals selected manually (marked with asterisks in Fig. 1-6), the COMPASS program automatically generated the sequences of retro-reactions based on the FGI-, C- and the DIA-rules. Since up to 20 precursors can be generated for each subgoal, the fully automatic generation of the entire "synthesis tree" was obviously impossible for the multistage computer search.

The reproduced total syntheses have been shown in this paper in a traditional manner as the retro-synthetic route from the target compound to the initial structures. All steps are traditionally described as antithetic transformations. The type of applied DIA-rule (F-σ-, R-σ- or π-approximation) is placed over the corresponding arrow. Similarly, the brief notation of the M- and the C-rules is specified for the corresponding retro- reaction.

The scheme for the first published synthesis of quadrone,[30] reproduced by the COMPASS program, is shown in Fig.1. Almost all retro- synthetic steps are described as F-σ-approximations (lactonization, condensation, alkylation, Michael addition). There is only one step where COMPASS cannot apply the DIA-rule directly and applies the M-rule, interpreted as an oxidation of a halogenalkyl substituent to a carbonyl group. This modification leads to the appropriate displacement of activated centers, and the resulting structure fragment appears appropriate for the direct application of the DIA-rule. In turn, the latter can be interpreted as a retro-Michael addition (see the last F-σ-approximation in Fig.1).

The next total synthesis of quadrone[31] reproduced by the COMPASS program, is presented in Fig.2. It should be noted that the authors[31] describe the synthesis not just of quadrone, but of its precursor **10**. It is easy

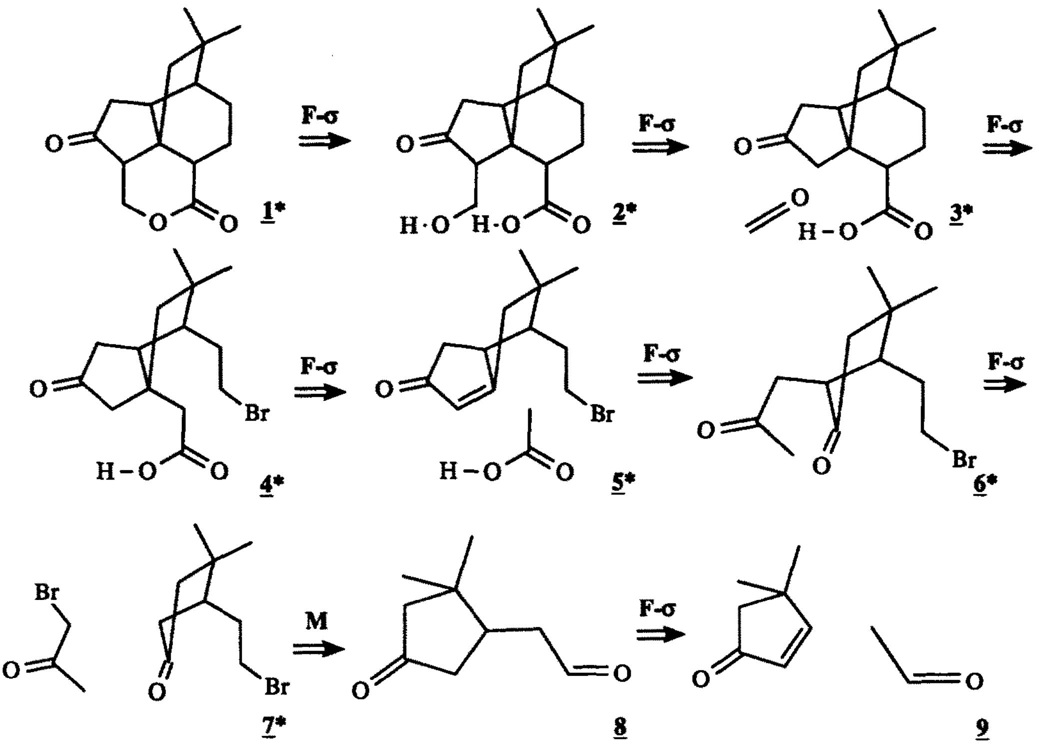


Fig. 1. Reproduction of the total synthesis of quadrone[30] by the COMPASS program

to see that structure **10** can be found in the quadrone "synthesis tree" as a result of the application of an appropriate M-rule (introduction of a double bond to the skeleton) to the structure **3** in Fig. 1. Therefore, the retro-synthetic steps from quadrone to **10** are omitted in Fig. 2.

As in the previous retro-synthesis, all stages are described in terms of F-σ-approximations (condensation, Grignard synthesis, alkylation). The two-stage application of the M-rules leads to the possibility of carrying out the retro-addition reaction, and this step is in complete agreement with the retro-analysis of the real synthesis.

Another branch of the "synthesis tree" generated by the COMPASS program for quadrone, is presented in Fig.3. As in the previous case, the retro-synthetic sequence from quadrone to **10** has been omitted for brevity (see also Fig.1). The key stage in the real synthesis[32] is the elegant implementation of a Wagner-Meerwein rearrangement. It is shown in Fig.3 that the COMPASS program reproduces this rearrangement as a result of applying the DIA-rule from the set of R-σ-approximations.

It is interesting to note further how the program examines the possibility of applying the DIA-rule, which corresponds to a retro- Grignard synthesis in structure **19**. The direct implementation of the appropriate DIA-rule is impossible due to the presence of the carbonyl group in **19**. Therefore, the COMPASS program uses the M-rule in order to change the carbonyl group (diactivated center in terms of the DIA-approach) by a triple bond. The corresponding oxidation of the triple bond to the carbonyl group is well-known in organic chemistry. It should be stressed that there is no other information concerning this reaction in the COMPASS knowledge base. Nevertheless, the generalized character of the rules applied makes it possible to reproduce this real synthetic method completely.

Let us consider the remaining retro-synthetic steps in Fig.3. The application of the DIA-rule

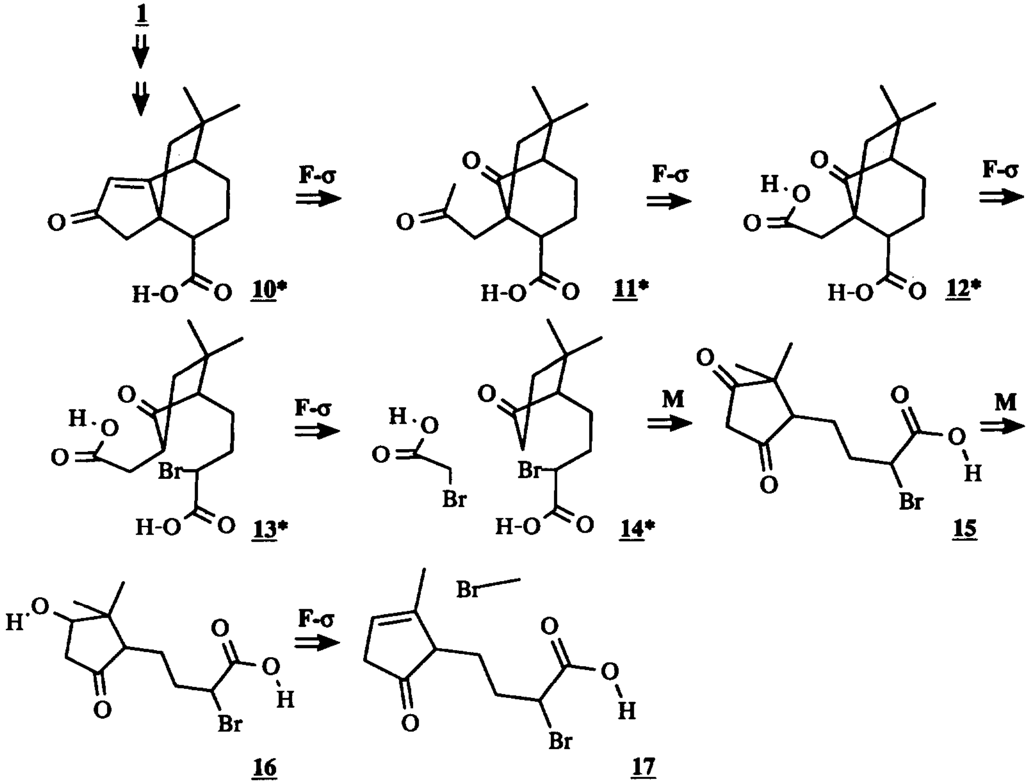


Fig.2. Reproduction of the total synthesis of quadrone[31] by the COMPASS program

(F-σ-approximation) can be interpreted as a retro-acylation of the cyclohexanone structure (transformation 22 → 23). The transformation 23 → 24 is an example of a π-approximation for two bonds, dissected from a 4-member ring with the formation of two simple quadrone precursors. It should be emphasized that the two compounds 24 were chosen as starting materials in an actual total synthesis of quadrone[32].

Fig.4 shows a part of the total quadrone synthesis generated. In this synthesis the key stages correspond closely to those of the published total synthesis.[33] At the first stage of the retro-synthesis the lactone cycle cleavage is generated. It should be taken into account that the COMPASS program carries out the simultaneous dissection of C-O bonds in the lactone cycle. These dissections are also followed by simultaneous F-σ-approximations. Such "parallel" consideration of bonds at the individual heteroatom avoids the non-informational enumeration of variants with subsequent dissection of each carbon-heteroatom bond.

The next stage is the application of another DIA-rule (F-σ-approximation), which can be interpreted and visualized as a retro-aldol condensation. Further, the COMPASS program reproduces the intermediate complication of quadrone precursor skeleton 3. We note that the knowledge base in the COMPASS program has a rule for introducing a new carbon atom in a skeleton at the neighboring position to the diactivated center. Hence, if the set of precursors of structure 3 is generated with the application of C-rules, it must necessarily contain the precursor 25.

Two different M-rules are implemented in the retro-transformation 25 → 27: i) modification of the non-activated center to the diactivated one, interpreted for structure 25 as an oxidation of the sp^3-carbon atom in the cyclohexane ring; ii) modification of the diactivated center to a monoactivated one with the parallel

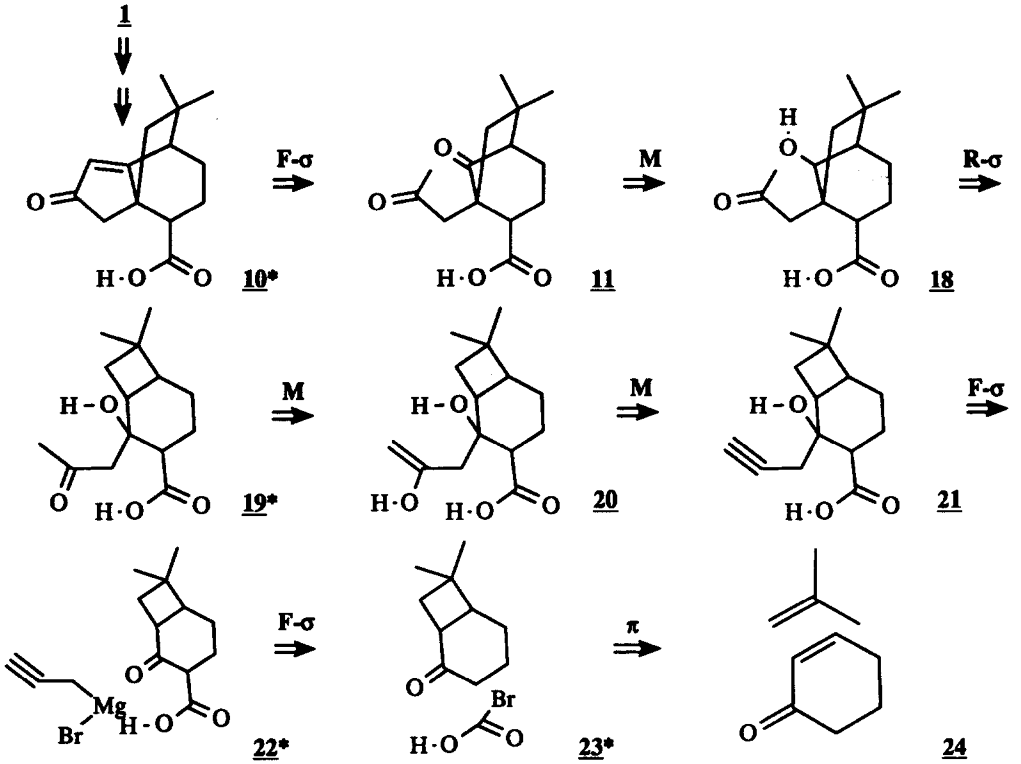


Fig.3. Reproduction of the total synthesis of quadrone[32] by the COMPASS program

formation of a double bond, traditionally interpreted as a keto-enol tautomeric transformation.

The application of these M-rules leads to the specific structural fragment which is suitable for the direct application of the extremely effective R-σ-approximation (Claisen retro-condensation for 27 in Fig.4). We note here that the Claisen rearrangement is a key stage in this total synthesis of quadrone[33], and the COMPASS program reproduces this transformation in a 6-step retro-synthesis.

Among the remaining transformations in Fig.4, the example of a retro-complication of skeleton 35 → 36 particularly should be noted. The corresponding C-rule can be interpreted as a retro-ozonolysis and the resulting cyclohexene formation permits the application of the previously mentioned Claisen retro-condensation in structure 36.

Another variant of the total quadrone synthesis[34] is presented in Fig.5. This synthesis was carried out by the same research group[33] and therefore some of the duplicated stages are omitted. The key stage in the presented synthetic pathway, as well as in the practical synthesis,[33] is the application of the Claisen rearrangement, although the appropriate structure fragment is formed differently. Three retro-reactions in Fig.5 are of particular interest. For the first, the C-rule in transformation 2 → 40 which can be interpreted as a process of retro-ozonolysis with cyclopentene ring formation (compare with the transformation 35 → 36 in Fig.4). For the second, the application of an aldol retro-condensation in structure 41 followed by a two-step modification of the activated centers. For the third, the Claisen retro-rearrangement 27 → 28 reproduced by the COMPASS program as an application of the DIA-rule (R-σ-approximation of dissected bond).

In order to convince the reader that the COMPASS program is universal and not just applicable to the

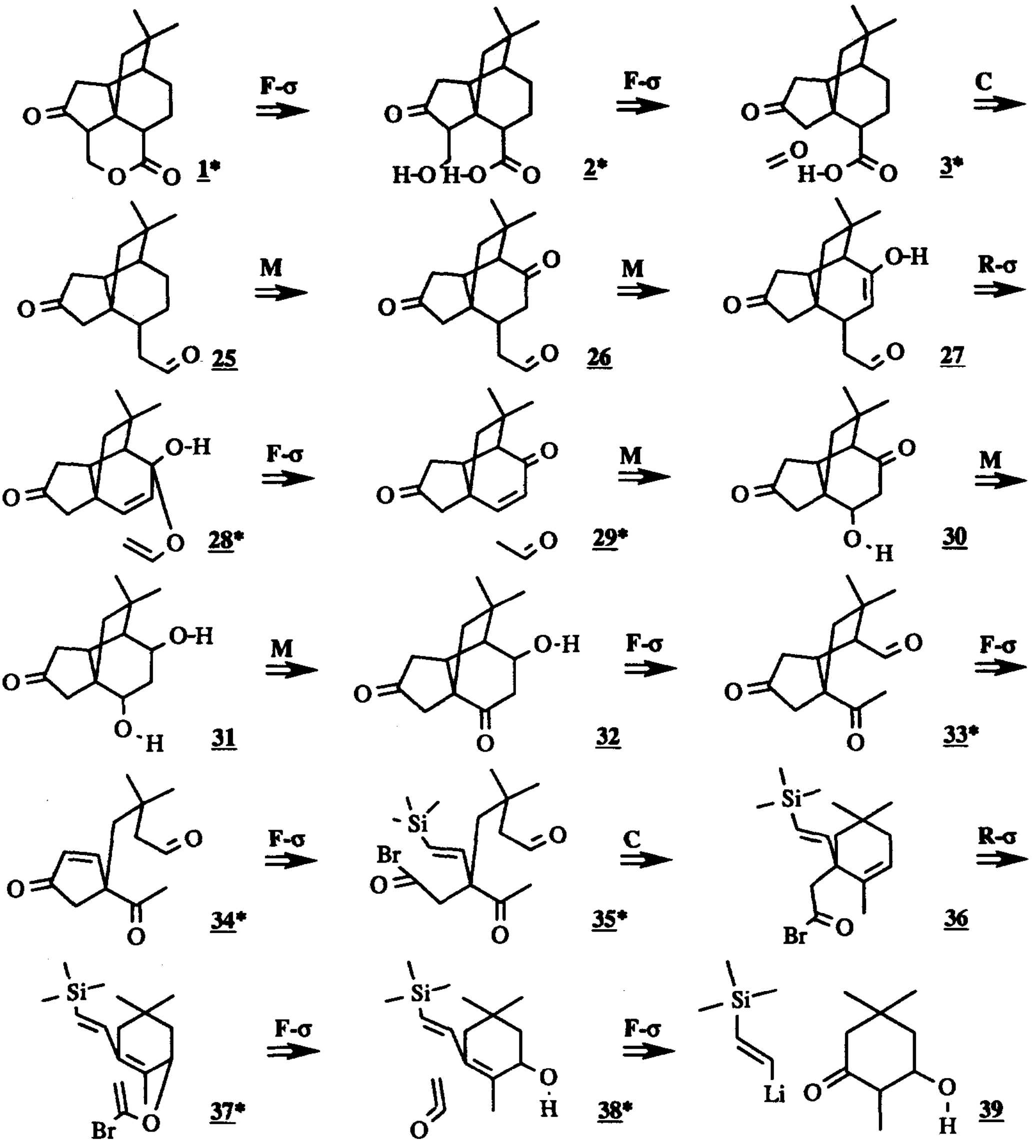


Fig.4. Reproduction of the total synthesis of quadrone[33] by the COMPASS program

reproduction of total quadrone syntheses, we also present an example of the retro-synthesis of sirenin[42]. It is easy to see in Fig.6 that the target structure is simpler than the quadrone structure, and the total synthesis includes only 6 steps. Although simple, it is an interesting demonstration of the reproduction of classical synthetic methods by the COMPASS program. Indeed, the first and the third stages in Fig. 6 (application of F-σ-approximations) can be interpreted as a reproduction of Wittig retro-reactions. The transformation 46 → 47 is also of interest, because the corresponding direct reaction is the well-known Favorskii rearrangement with ring contraction. The last retro-reaction 47 → 48 is an example of the π-approximation application and

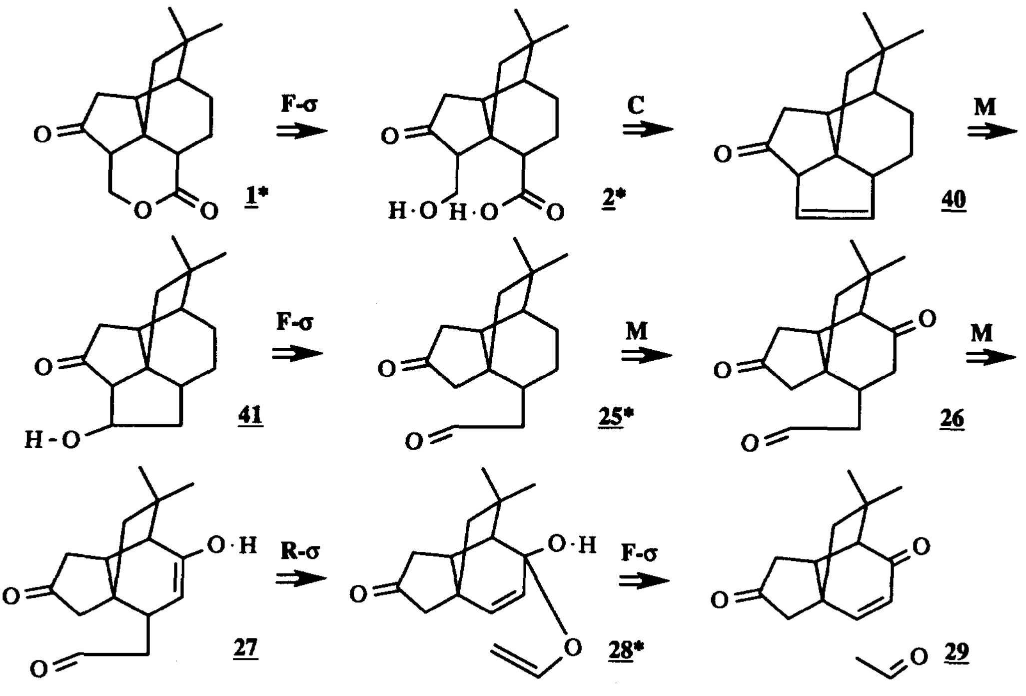


Fig.5. Reproduction of the total synthesis of quadrone[34] by the COMPASS program

can be interpreted as a [2+2] retro-cycloaddition.

The presence of a hydroxyl group in the ketene component in **48** seems unreasonable due to the instability of the structure. It should be noted that in the actual synthesis[42] a chlorine atom, not the hydroxyl group, is present in the ketene structure. This incompatibility between the literature synthesis and the retro-synthetic pathway generated by the COMPASS program occurs because of the generalized manner of the FG description in the DIA-approach. Actually, the halogen substituent and the hydroxyl group are not distinguished in the DIA-approach and the actual symbols of atoms are inserted in a structural formula in a special visualization module of the COMPASS program. This module prepares the output of information in the most traditional and conventional manner from a chemist's point of view. For instance, the generalized description of the F-σ-approximations in structures **42** and **44** is interpreted in the visualization module as a Wittig retro-reaction. Nevertheless, the example of the hydroxylketene in **48** shows that this module must be improved in order to avoid displaying essentially correct information, in an incorrect manner.

CONCLUSION

In this paper we have tried to demonstrate to chemists that the present COMPASS version is a complete and useful CAOS program. The results of program testing have convinced us that COMPASS can be used as a working tool for both planning routine organic synthesis and in the search for new promising synthetic pathways. The knowledge base used by the COMPASS program is universal and is not just limited in the planning of synthetic routes to natural compounds. Moreover, the flexible structure of the COMPASS program allows the additional extension of the knowledge base, if required. It is our belief that only a few rules will be added in order to describe some specific heterocyclic transformations.

The possibility of implementing the non-interactive mode of the COMPASS program operations also

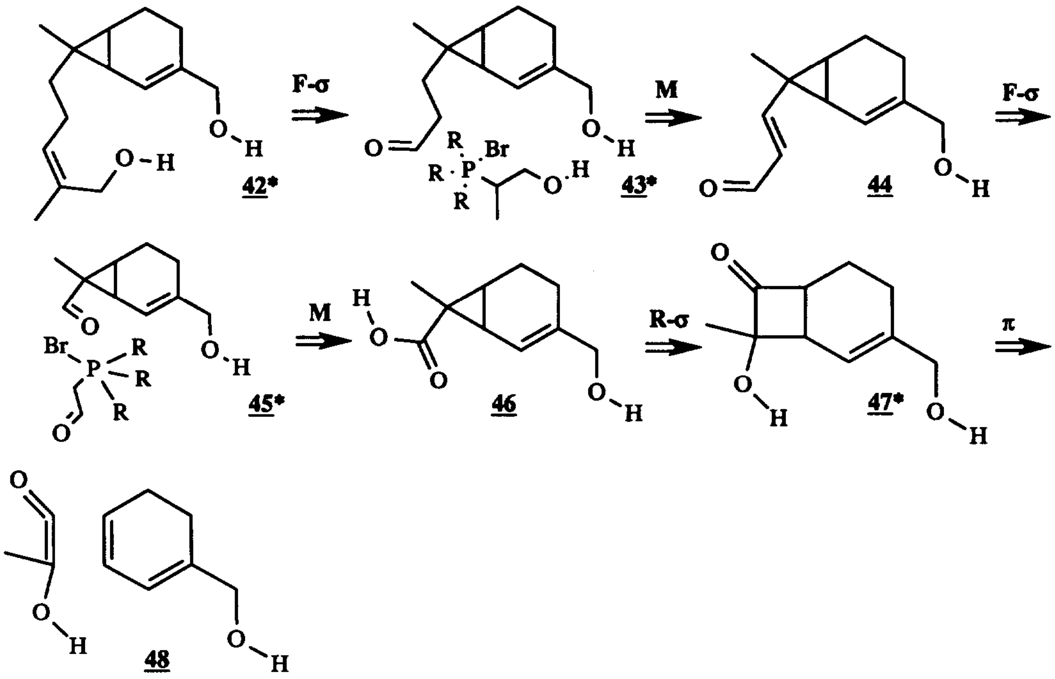


Fig.6. Reproduction of the total synthesis of sirenin[42] by the COMPASS program

appears to be very attractive. This mode should provide the user with a fully automatic selection of the appropriate dissections in the target structure. Thus, in traditional CAOS terminology, the problem is to create the selection criteria for choosing strategic bonds[43] in a structure. However, the CAOS history shows that there is no universal solution to this problem and any attempt to select strategic bonds without the chemist's participation should be tried with caution, otherwise the chance of losing valuable and interesting solutions increases. It should be remembered that the present version of the COMPASS program exhaustively generates the set of all possible dissections which can be accepted completely or partially depending on the user's wish. So far, we have tried to use only the most evident and indisputable criteria to select reactions and structures generated by a non-redundant combinatorial enumeration. Nevertheless, some promising measures have been already developed for molecular complexity[43-45] and we have plans to apply this non-empirical criterion for automatic structure selection.

ACKNOWLEDGEMENT

We are thankful to Dr. Richard Barcock (University of Florida, Gainesville) and Ms. Mary A. Austin (Duluth, Minnesota) for their kind assistance in checking the English.

REFERENCES

1. Bersohn, M.; Esack, A. *Chem.Rev.* **1976**, *76*, 269-282.
2. Gasteiger, J.; Hutchings, M.G.; Christoph,B.; Gann, L.; Hiller, C.; Low, P.; Marsili, M.; Saller, H.; Yuki, K. *Top. Curr. Chem.* **1987**, *137*, 19-73.
3. Barone, R.; Chanon, M., In *Computer Aids to Chemistry*; Vernin, G.; Chanon, M., Eds, E.Horwood: Chichester, **1986**, 19-102.

4. Zefirov, N.S.; Gordeeva, E.V. *Russ.Chem.Rev.* **1987**,*56*,1002-1014.
5. Corey, E.J.; Wipke, W.T. *Science* **1969**,*166*,178-192.
6. Wipke, W.T.; Braun, H.; Smith, G.; Choplin, F.; Sieber, W. *ACS Symp. Ser.* **1977**,*61*,97-127.
7. Corey, E.J.; Pensak, D.A. *ACS Symp.Ser.* **1977**,*61*,1-32.
8. Haggin, J. *Chem.Eng.News* **1983**,*61*,7-20.
9. Jochum, C.; Gasteiger, J.;Ugi, I.; Dugundji, J. *Z.Naturforsch.* **1982**, *37b*,1205-1215.
10. Gasteiger, J. *Tetrahedron* **1979**,*35*,1419-1426.
11. Hendrickson, J.B. *J.Am.Chem.Soc.* **1977**,*99*,5439-5450.
12. Zefirov, N.S.; Gordeeva, E.V.; Tratch, S.S. *J.Chem.Inf.Comp.Sci.* **1988**, *28*,188.
13. Zefirov, N.S.; Gordeeva, E.V. *J.Org.Chem.* **1988**,*53*,527-532.
14. Brandt, J.; Friedrich, J.; Gasteiger, J.; Jochum, C.; Schubert, W.; Ugi, I. *ACS Symp.Ser.* **1977**,*61*,33-59.
15. Gasteiger, J.; Jochum, C. *Top.Curr.Chem.* **1978**,*74*,93-126.
16. Gordeeva, E.V.; Shcherbukhin, V.V.; Zefirov, N.S., *accepted in Tetrahedron Computer Methodology.*
17. Gasteiger, J. *Comput.Chem.* **1978**,*2*,85-88.
18. Weise, A. *Z.Chem.* **1975**,*15*,333-340.
19. Doenges, R.; Groebel, B.T.; Nickelsen,H.; Sander,J. *J.Chem.Inf.Comput.Sci.* **1985**,*25*,425-430.
20. Bures, M.G.; Roos-Kozel, B.L.; Jorgensen, W.L. *J.Org.Chem.* **1985**,*50*, 4490-4498.
21. [a]Metivier, P.; Gushurst, A.J.; Jorgensen, W.L. *J.Org.Chem.* **1987**,*52*, 3724-3738; [b]Jorgensen, W.L., Laird, E.R., Gushurst, A.J., Fleisher, J.M., Gothe, S.A., Helson, H.E., Paderes, G.D., Sinclair, S. *Pure Appl.Chem.* **1990**,*32*, 1921-1932.
22. Azario, P.; Barone, R.; Chanon, M. *J.Org.Chem.* **1988**,*53*,720-724.
23. Mehta, G.; Azario, P.; Barone, R.; Chanon, M. *Tetrahedron* **1989**, *45*,1985-1994.
24. Barone, R.; Chanon, M. *Chimia* **1986**,*40*,436-439.
25. Hendrickson, J.B.; Grier, D.L.; Toczko, A.G. *J.Am.Chem.Soc.* **1985**,*107*, 5228-5238.
26. [a]Hendrickson, J.B.; Toczko, A.G. *Pure Appl.Chem.* **1988**,*60*,1563-1572; [b]Hendrickson, J.B., Toczko, A.G. *J.Chem.Inf.Comp.Sci.* **1989**,*29*, 137-145; [c]Hendrickson, J.B., Miller, T.M. *J.Am.Chem.Soc.* **1991**, *113*, 902-910.
27. Schlessinger, R.H.; Wood, J.L.; Poss, A.J.; Nugent, R.A.; Parsons, W.H. *J.Org.Chem.* **1983**,*48*,1146-1147.
28. Mehta, G.; Reddy, A.V. *J.Chem.Soc.Chem.Commun.* **1981**,*N 15*,756-757.
29. Stork, G., Clark, G., Shirer, C.S. *J.Am.Chem.Soc.* **1981**,*103*,4948-4949.
30. Danishefsky, S.; Vaughan, K.; Gadwood, R.; Tsuzuki, K. *J.Am.Chem.Soc.* **1981**, *103*,4136-4141.
31. Kende, A.S.; Roth, B.; Sanfilippo, P.J.; Blacklock,T.J. *J.Am.Chem.Soc.* **1982**, *104*,5808-5810.
32. Takeda, K.; Shimoto, Y.; Yoshii, E. *J.Am.Chem.Soc.* **1984**,*105*,563-568.
33. Burke, S.D.; Murtiashaw, C.W.; Saunders, J.O.; Dike, M.S. *J.Am.Chem.Soc.* **1982**,*104*,872-874.
34. Burke, S.D.; Murtiashaw, C.W.; Saunders, J.O.; Oplinger, J.A.; Dike, M.S. *J.Am.Chem.Soc.* **1984**, *106*,4558-4566.
35. Bornack, W.K.; Bhagwat, S.S.; Ponton, J.; Helquist, P. *J.Am.Chem.Soc.* **1981**, *103*,4647-4648.
36. Dewanckele J.M.; Zimmerman, F.; Vandewalle, M. *Tetrahedron* **1983**,*39*, 3235-3244.
37. Piers, E.; Moss, N. *Tetrahedron Lett.* **1985**,*26*,2735-2738.
38. Smith,III, A.B.; Konopelski, J.P. *J.Org.Chem.* **1984**,*49*,4094-4095.
39. Wender, P.A.; Wolanin, D.J. *J.Org.Chem.* **1985**,*50*,4418-4420.
40. Funk, R.L.; Abelman, M.M. *J.Org.Chem.* **1986**,*51*,3247-3248.
41. Liu, H.J.; Brunet, M.L. *Can.J.Chem.* **1988**,*66*,528-530.
42. Harding, K.E.; Strickland, J.B.; Pommerville, J. *J.Org.Chem.* **1988**,*53*, 4877-4883.
43. Corey, E.J.; Howe W.J.; Orf, H.W.; Pensak, D.A.; Petersson, G. *J.Am.Chem.Soc.* **1975**,*97*,6116-6124.
44. Bertz, S.H. *Bull.Math.Biol.* **1983**, *45*,849-855.
45. Bonchev, D; Polansky, O.E. In *Studies in Physical and Theoretical Chemistry*; Elsevier: Amsterdam, **1987**,*51*,126-158.