COMPUTER-ASSISTED DESIGN OF SYNTHESES FOR HETEROCYCLIC COMPOUNDS

Ralf Fick, Wolf-Dietrich Ihlenfeldt, and Johann Gasteiger*

Computer-Chemie-Centrum, University Erlangen-Nuernberg, Naegelsbachstr. 25, D-91052 Erlangen, Germany

Abstract - Fundamental strategies and search methods for the development of a synthesis plan for heterocyclic compounds are presented. These methods have been incorporated into the synthesis design system WODCA (Workbench for the Organization of Data for Chemical Applications). The use of this computer system is illustrated with several target compounds containing heterocycles of various degrees of complexity such as lysergic acid, 3-benzylpiperidine, and rotenone.

INTRODUCTION

A cornucopia of methods has been developed for the synthesis of heterocyclic compounds. Sometimes it appears that each different heterocyclic system asks for synthesis methods of its own. Because of this large variety of approaches to the synthesis of heterocyclic compounds the computer-assisted design of such syntheses offers particular challenges.

The concept of retrosynthesis as pioneered by Corey [1] has profoundly influenced the planning of organic synthesis. A stepwise retrosynthesis approach has been the cornerstone of most computer systems for the design of organic syntheses. However, it has to be realized that the various sequences of retrosynthetic steps that are generated and have to be evaluated lead to a scheme, the so-called synthesis tree, that rapidly becomes quite complex for anything but the simplest organic molecules (Figure 1).

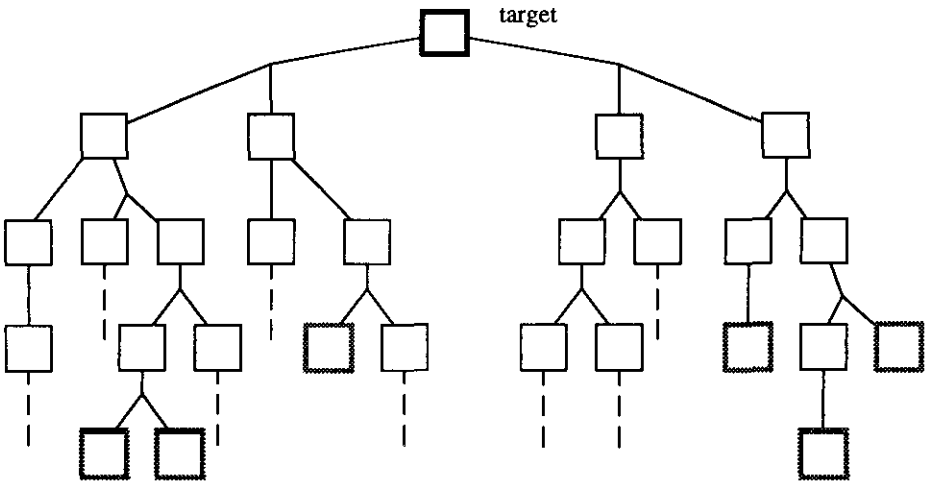

Figure 1 A synthesis tree as obtained from a stepwise retrosynthetic analysis showing the sequences of synthesis steps and - as squares - the target molecule, the synthesis precursors, and starting materials.

It is therefore increasingly felt that a variety of methods above and beyond a stepwise retrosynthetic approach must be included in a synthesis design system to efficiently assist the chemist in developing a synthesis plan and find a pathway to available starting materials as rapidly as possible.
In recognition of this need we have developed the WODCA (Workbench for the Organization of Data for Chemical Applications) system that offers a broad spectrum of search methods to find synthesis precursors, available starting materials and reactions to perform the various steps of a synthesis. [2-4]
In particular, one of the most challenging problems is a search for promising starting materials directly from the given target compound. For such search methods that cut across in many individual reaction steps allow one to develop efficient synthesis strategies and to drastically reduce the complexity of the synthesis tree (Figure 2).

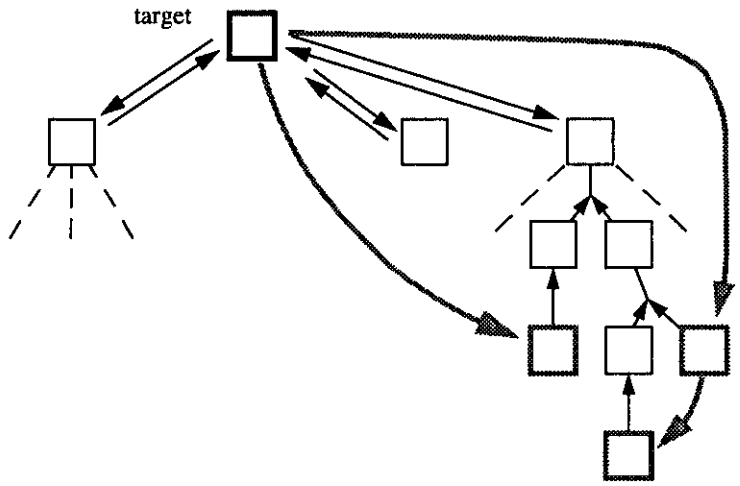

Figure 2 A synthesis tree narrowed by a variety of search methods that include the perception of similarity between the target and available starting materials, retrosynthetic steps and reaction prediction steps, as well as forward planning.

In this paper we will concentrate on such direct searches for promising starting materials in catalogs of available compounds. Where such searches directly from the target molecule do not produce good answers, the target compound will be converted to synthesis precursors by searching for strategic bonds and then breaking these bonds to generate precursor molecules. These will then be used for their side to search for available starting materials.

STARTING MATERIALS FOR THE SYNTHESIS OF HETEROCYCLIC COMPOUNDS

The search for starting materials for a given target molecule centers around novel definitions of chemical similarity, similarity criteria that have been developed with a particular eye on chemical synthesis. These similarity criteria fall into two categories, those based on the occurrence of large substructures common both to the target (search) molecule and a starting material (hit) and those similarity measures that are based on general reaction types. [3]
To give an example for both types:
Substructure based: Two compounds are considered similar if they contain the same ring system (e.g. an indole ring).
Reaction based: Two compounds are taken as similar if they can be interconverted by an oxidation or reduction reaction.

These similarity definitions have been translated into structural transformations that allow one to process an entire database of available starting materials prior to their use in a synthesis design system. [2,3] This makes the similarity search in a database of available starting materials extremely fast.

About 50 different similarity measures have been developed. Some have a wide range of applicability whereas others are rather specific albeit quite useful in special situations. Some of these measures have been coined for certain classes of compounds, others should be explored when everything else fails.

Several types of similarity measures appear to be particularly suited for the design of syntheses of heterocyclic compounds:

1. Those that ask for a common ring system in the target and starting material (Figure 3a).
2. Those that involve a reduction (or an oxidation) of a ring in the starting material to obtain the one in the target molecule (Figure 3b).
3. Those that require the target and the starting material to have a carbon chain with the same substitution pattern (Figure 3c).

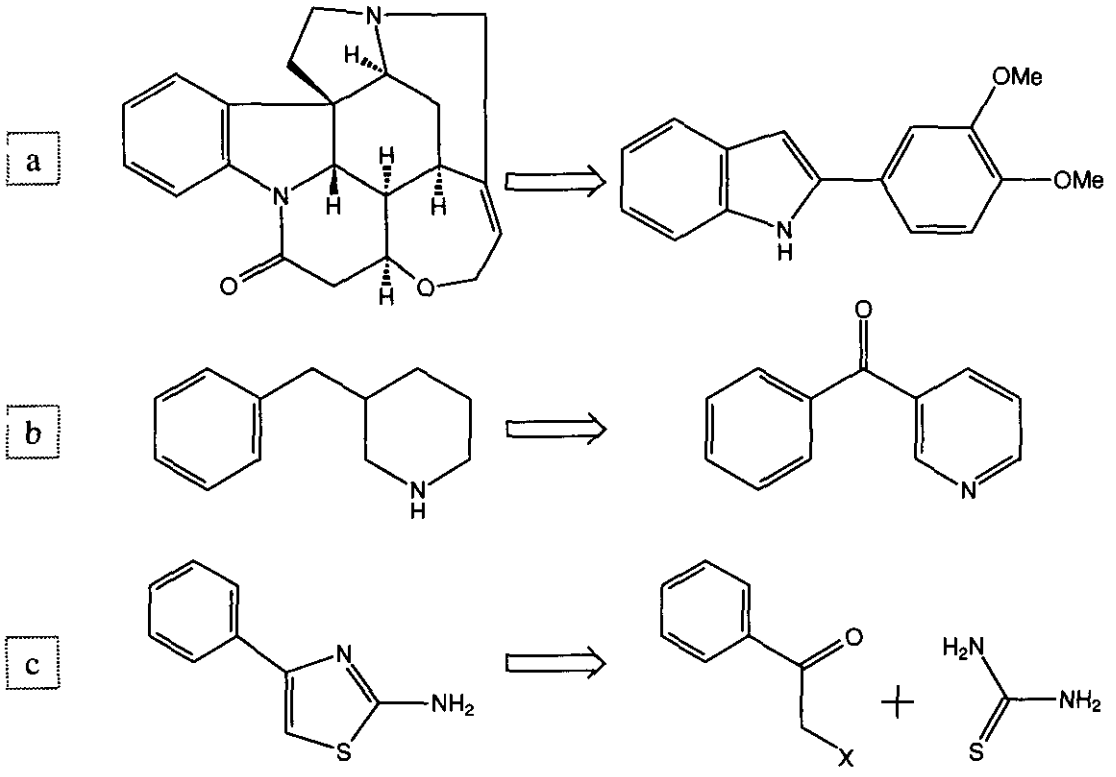

Figure 3 Basic strategies for the synthesis of heterocyclic compounds: a) Searching for a common ring system; b) oxidation / reduction of the ring system; c) construction of the heterocyclic system.

In this paper we will explore the usefulness of some of these similarity measures for the design of heterocyclic compounds by giving several examples of synthesis design studies.

As catalog of available starting materials the Janssen Chimica catalog containing 7,849 different compounds was used in most cases. For the synthesis of compounds having a specific stereochemistry, a catalog of 2,211 molecules (chiral pool compounds) compiled for the CHIRON program [5] is also incorporated into WODCA and was searched in a few cases.

LYSERGIC ACID

Lysergic acid (**1**) (Figure 4) has a fairly complex ring system and one cannot expect to find a commonly available starting material in a catalog of less than 8,000 compounds. Indeed none of the 50 similarity measures provided a hit. On the other hand, it is clearly a viable basis for the development of a synthesis for lysergic acid to start with a compound that already contains the indol ring system. 85 compounds were found in the Janssen Chimica catalog which have an indol ring. To narrow the number of precursors to such compounds that have a closer similarity to the target structure lysergic acid, more stringent structural requirements were set: The starting material was required to have a substituent at position 4 of the indol ring that is also substituted in lysergic acid. Breaking the bond at a position of the indol ring in lysergic acid gives structure (**2**) as synthesis precursor. This structure is used as the search query.

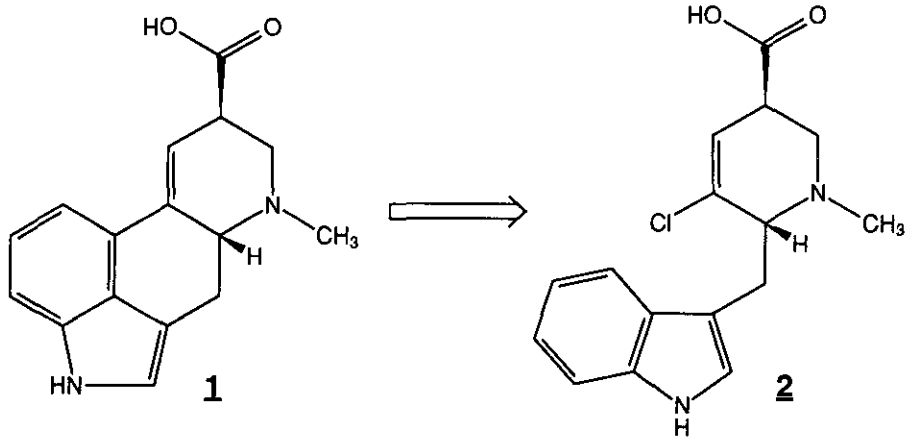

Figure 4 The target compound, lysergic acid (**1**) and a potential precursor (**2**).

A search was made with the similarity measure "ozonolysis and substitution pattern", that basically breaks CC-double bonds and requires the presence of any heteroatom at positions that already have heteroatoms in the precursor (**2**). This search provided 15 molecules from the starting material library that are shown in Figure 5.
To facilitate the planning of a synthesis these compounds are further evaluated and ranked by algorithms built into the WODCA system. This evaluation takes into account the stereochemistry of the search structure (**2**) and that of the starting material, as well as the amount of refunctionalization necessary to convert the starting material into the target structure.
All compounds perceived as promising starting materials are tryptophan derivatives. The highest merit value was given to D-(+)-tryptophan (**3**) having the same stereochemistry as required in lysergic acid. L-(-)-tryptophan (**5**) is less favourably evaluated as an inversion of stereochemistry would be necessary. L-(-)-abrine (**4**) has the additionally required methyl substituent on the amino group and therefore obtains a slightly better evaluation than L-(-)-tryptophan. The racemate DL-tryptophan (**6**) obtains a lower merit value than the amino acid with the wrong stereochemistry. The rationale behind this evaluation is that an inversion of a stereochemical center (by a S_N2 reaction) is easier to perform than to use a racemate which asks for a racemate separation with concomitant loss of half of the starting material or the requirement for an additional stereochemical inversion operation after racemate separation.
This brief report shows how the use of basic search mechanisms in WODCA - the definition of strategic bonds and similarity searches in a database of starting materials - allows the user to rapidly find a promising starting material for the synthesis of lysergic acid. Clearly, quite some more work has to be done to develop a full plan for the synthesis of lysergic acid. However, with both the target and a fairly large starting material being defined, the search can now be narrowed to a much smaller part of a synthesis tree.

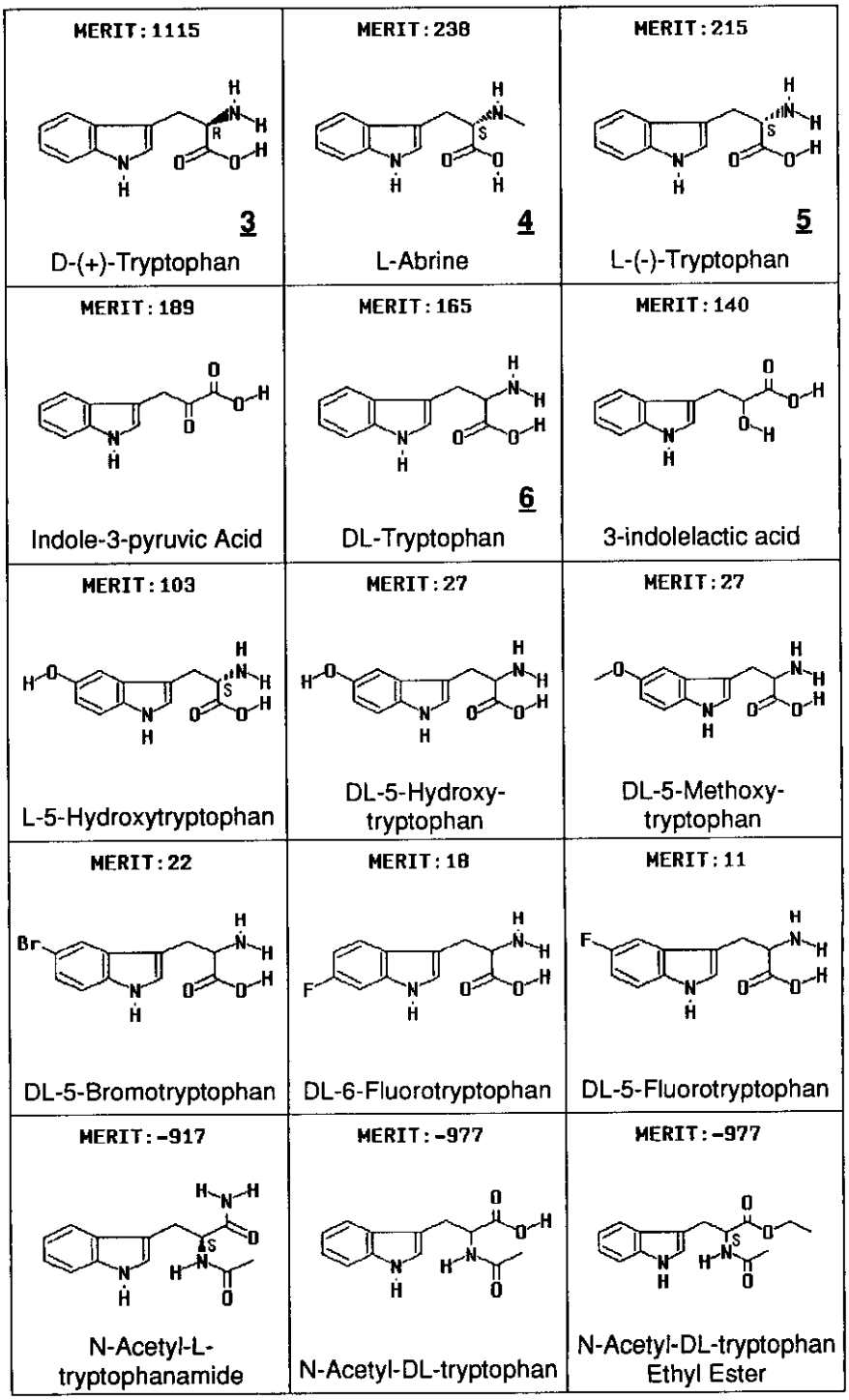

Figure 5 Starting materials found to be similar to the precursor molecule (**2**).

3-BENZYLPIPERIDINE

Whereas lysergic acid comprised a target with a rather complex structure, the next molecule to be synthesized, 3-benzylpiperidine (**7**) has quite a simple structure (Figure 6).

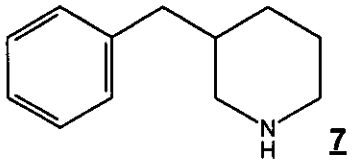

Figure 6 The target molecule, 3-benzylpiperidine (**7**).

Nevertheless, this example nicely illustrates the interplay between different planning methods in synthesis design.
With a simple structure like that given by 3-benzylpiperidine a similarity search in a catalog of available starting materials with the target being the direct query seems justified. Syntheses of saturated heterocyclic compounds by hydrogenation of heteroaromatic systems and vice versa are quite common. This idea is embedded in the similarity criterion "carbon skeleton with complete reduction" that allows complete hydrogenation of heterocyclic systems. A search based on this similarity definition with 3-benzylpiperidine in the Janssen Chimica catalog provides the five structures shown in Figure 7.

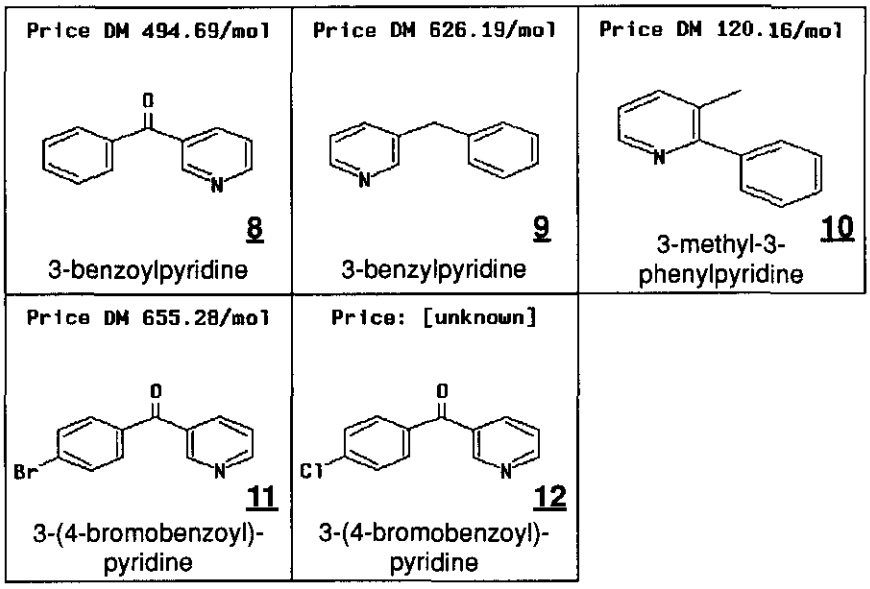

Figure 7 Starting materials found to be similar to 3-benzylpiperidine (**7**) by the transformation "carbon skeleton with total reduction" (observe the price of the starting materials).

For four of the five structures (**8**, **9**, **11** and **12**) it is quite clear that they can be converted into the target by reduction. But why was compound (**10**), 3-methyl-2-phenylpyridine found? The structural transformation defining the similarity criterion "carbon skeleton with complete reduction" converts 3-benzylpiperidine to the basic search structure (**13**) (Figure 8). This skeleton is also obtained from all the structures shown in Figure 7, including 3-methyl-2-phenylpyridine (**10**).

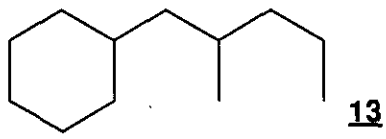

Figure 8 The skeleton that forms the basis of similarity of the structures (7 - 12) (Figures 6 and 7).

In fact, one can conceive a pathway from 3-methyl-2-phenylpyridine (10) to 3-benzylpiperidine: Bromination of the methyl group, hydrogenation of the pyridine ring, ring opening and then ring closure to the bromomethyl group, and, finally, reduction of the benzylic position. Clearly, this sequence is less attractive than those that come to mind with the other four starting materials (8, 9, 11 and 12). *Nevertheless, it shows that the similarity* found by this criterion can be put to work also in situations where there are not so attractive alternatives.
Having found commercially available starting materials one could stop the synthesis search. However, if one finds the price of the starting materials too high (prices are indicated in Figure 7) the synthesis search can be driven to simpler and less expensive starting materials.
Let us do that with 3-benzoylpyridine (8) a compound that can easily be converted into the target compound (7) by hydrogenation over RaneyNi. [6]
A search for strategic bonds indicates - not surprisingly - the two bonds to the carbonyl groups to be strategic. Breaking these two bonds to the carbonyl group and converting the synthons thus obtained into reagents provides the two alternative pathways shown in Figure 9.

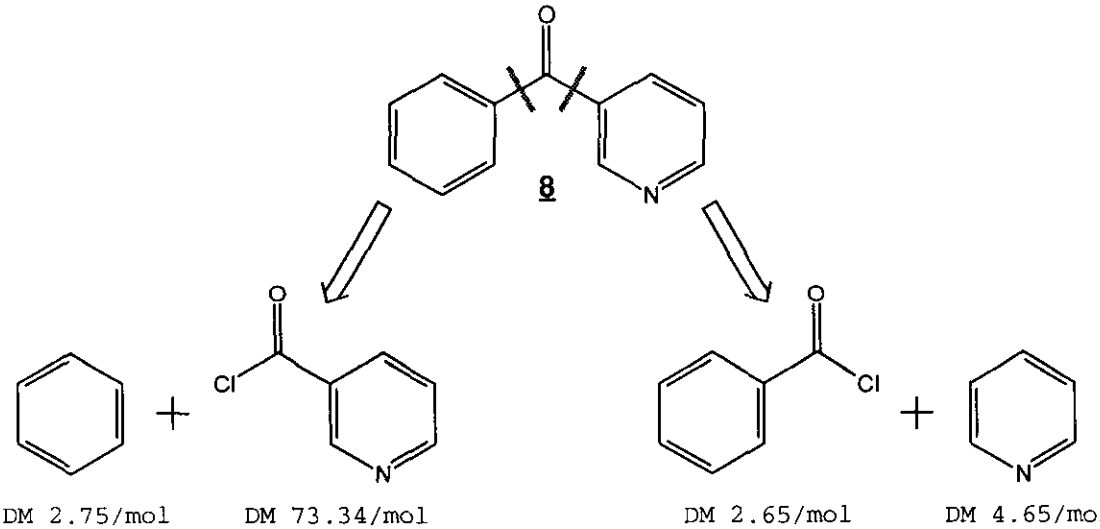

Figure 9 Two approaches of the synthesis of 3-benzoylpyridine (8), obtained on the basis of the automatic *detection of strategic bonds* (observe the price of the starting materials).

The prices of the starting materials are much lower than those shown in Figure 7. This indicates that it is worth considering to add an extra step to the synthesis of 3-benzylpiperidine and start with simpler - and quite cheaper - starting materials. Furthermore, based only on the price of the starting materials - disregarding reaction times, yields, etc. - the synthesis that starts with benzoylchloride and pyridine is much more attractive.
We stay with this target structure, 3-benzylpiperidine (7) for a while to show other approaches to its synthesis. The recipe to success in the above synthesis study was our initial use of the similarity criterion "carbon skeleton

and complete reduction". Had we not used this similarity definition provided by WODCA, a search for strategic bonds might have seemed warranted. The procedures for the search and the evaluation of strategic bonds designed for carbon-heteroatom bonds and for bonds to aromatic rings give the result shown in Figure 10.

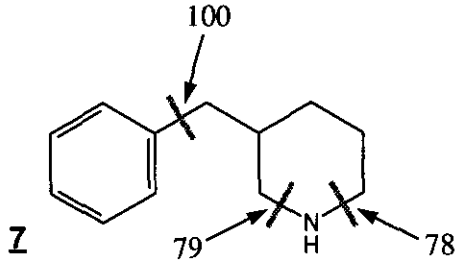

Figure 10 Strategic bonds in 3-benzylpiperidine and their scaled evaluation.

By choosing one such bond, WODCA provides the precursor (**14**). This compound is searched for its side for strategic bonds and the one with the highest evaluation is broken, giving benzene and 1,5-diamino-2-(chloromethyl)pentane (**15**) (Figure 11).

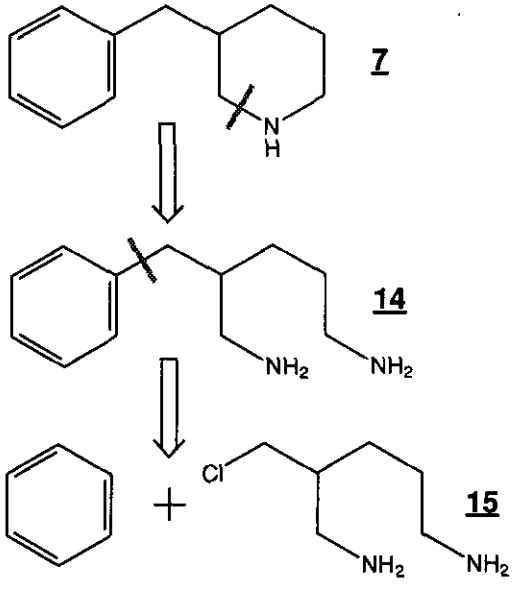

Figure 11 Retrosynthetic sequence developed for the synthesis of (**7**) by performing strategic bond evaluation, disconnection of such bonds, and precursor generation.

A similarity search for compounds having the same carbon skeleton and substitution pattern as (**15**) provides the twelve compounds shown in Figure 12 as available starting materials.
A quick scan of these structures indicates that only 2-cyanoethylmalonic ester (**16**) is an open chain compound. Continuing the synthesis search with this compound (**16**) leads to malonic ester (**17**) and acrylonitrile (**18**), inexpensive starting materials giving (**16**) in a Michael addition. This completes an alternative attractive approach to 3-benzylpiperidine. Figure 13 collects this alternative synthesis developed for the target molecule (**7**).

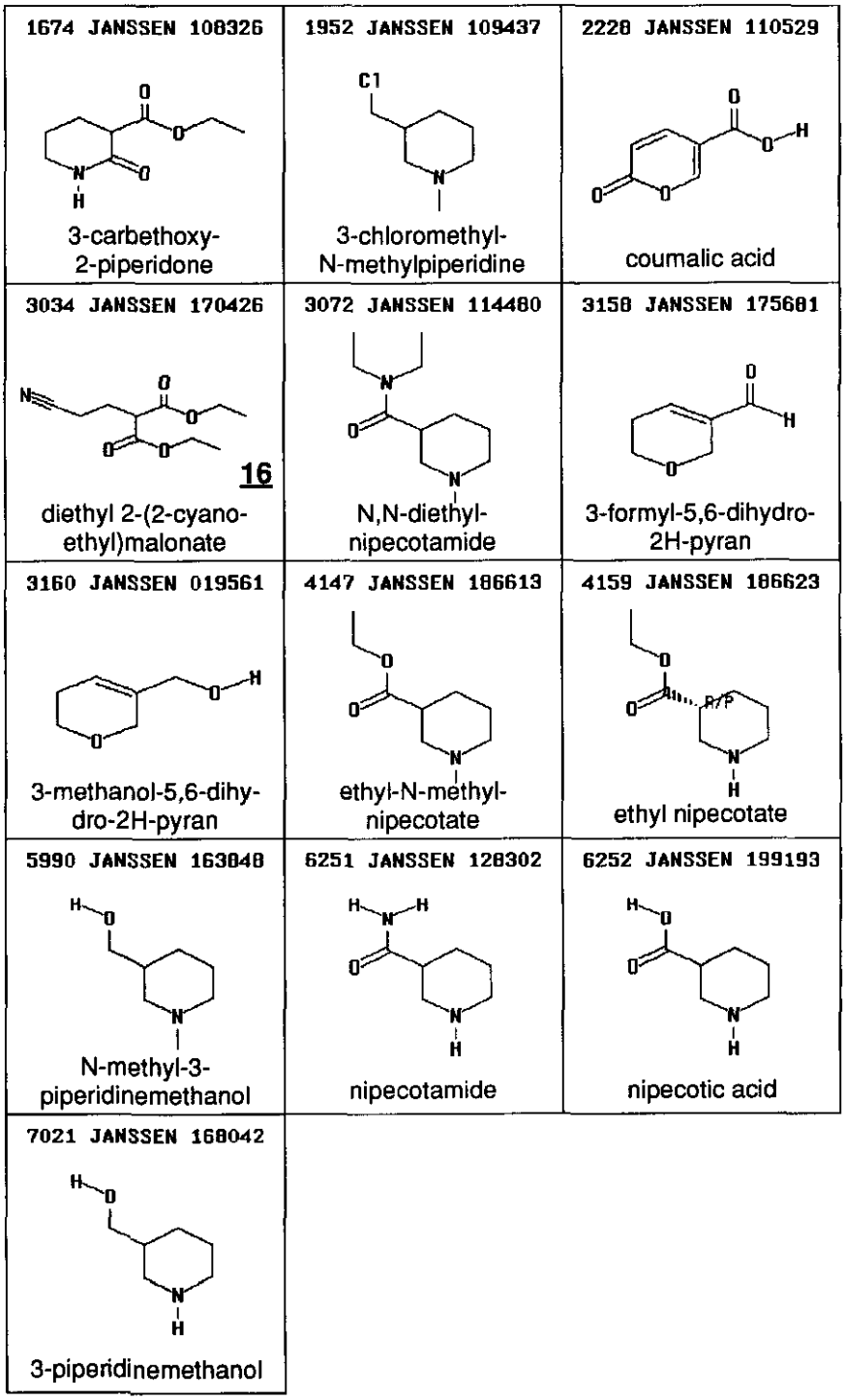

Figure 12 Compounds having the same carbon skeleton and substitution pattern as **15**.

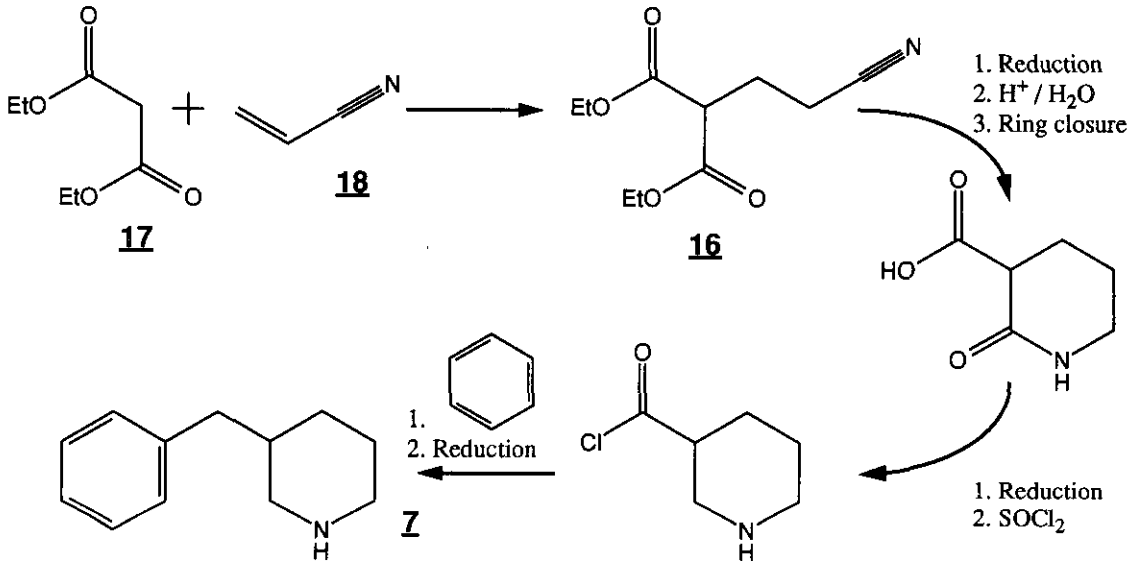

Figure 13 A third synthesis plan for 3-benzylpiperidine (**7**).

ROTENONE

Rotenone (**19**) (Figure 14) has quite a complicated ring system and therefore a simplification of the synthesis problem by converting it to precursor molecules appears to be the method of choice.

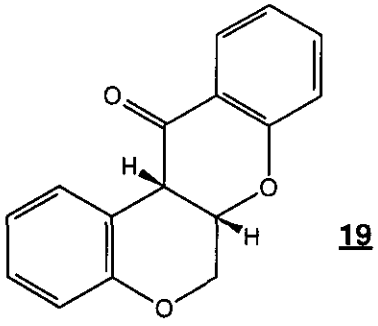

Figure 14 The rotenone skeleton (**19**).

Nevertheless, a similarity search was initiated with the entire target structure as query. To our surprise the similarity measure that searched for compounds having the same carbon skeleton but allowing also reductive interconversions ("carbon skeleton with complete reduction") provided a hit. The structure found is 2-carbethoxy-5,7-dihydroxy-4'-methoxyisoflavone (**20**) and is shown in Figure 15. This structure might be an indication of the biosynthesis of rotenone.[7] However, this pathway is not further explored here.

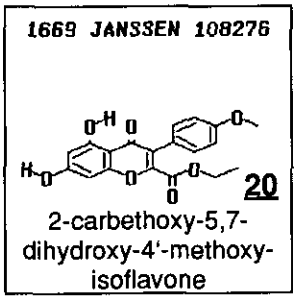

Figure 15 Structure (**20**) found to be similar to rotenone (**19**) with the criterion "carbon skeleton and complete reduction".

One of the problems that have to be addressed in the design of a synthesis for rotenone is the stereochemistry at the junction of the two heterocyclic rings. We will first search for a synthesis of the rotenone system disregarding the stereochemical problem and give it considerations only after having found a route deemed attractive.
Clearly, there are many routes possible for the synthesis of the rotenone system. We will not try to exhaustively explore the synthesis pathways to this structure nor will we strive to find the "best" route (whatever this means). Rather, we will show here only two pathways out of several that we have developed. In fact, the first pathway will only be hinted at and not explored in detail.

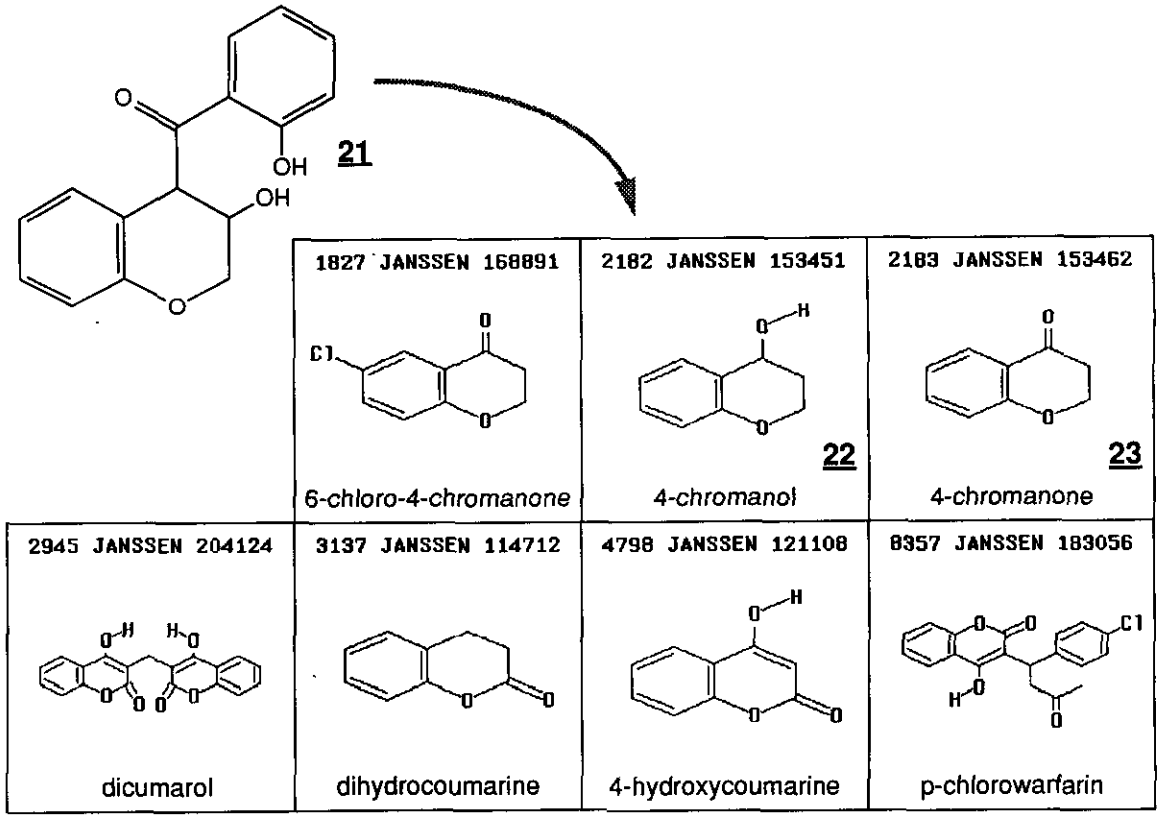

Figure 16 Available starting materials for the chromane ring system.

Several strategic bonds were found in rotenone with two of the carbon heteroatom bonds being the most prominent ones. Breaking the first bond provides compound (**21**). A search for compounds having this ring system, i.e. the larger chromane system, finds seven compounds in the Janssen Chimica catalog (Figure 16).
In fact, a synthesis for rotenone starting from **23** has already been developed. [8] The number of reaction steps from **23** to **21** amounted to eight. This underlines the potential of similarity searches to lay the foundation to long-range strategies.
The next synthesis route was developed by breaking the other highly evaluated carbon-heteroatom bond. This leads to the precursor (**24**) (Figure 17). This structure is investigated for its side for strategic bonds and here we make use of a feature of WODCA that allows one to break two bonds simultaneously. Breaking the two bonds indicated in **24** leads to precursor (**25**) and phenol (Figure 17).

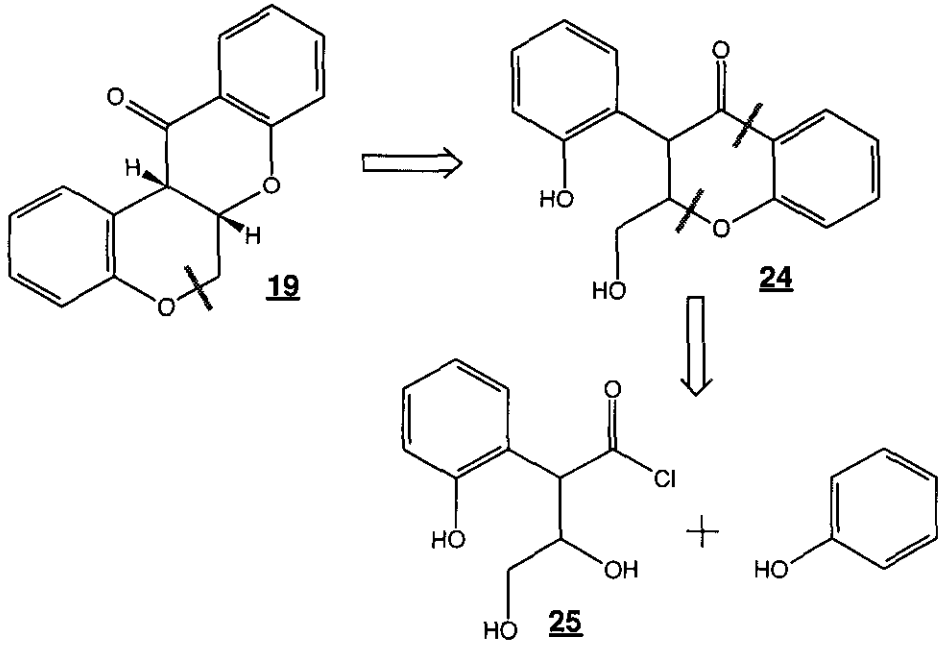

Figure 17 Second retrosynthesis for rotenone (**19**).

To investigate an access to precursor (**25**), a rather broad similarity search for compounds having the same carbon skeleton was performed. 45 compounds comprising quite a variety of structures, with the carbon atoms in different oxidation states, were found. A quick scan of these structures indicated several representatives of three different heterocyclic ring systems. An example for each of these three ring systems, an indole (**26**), a cumarine (**27**), and a quinoline derivative (**28**) is shown in Figure 18.
The structural similarity of these three compounds (**26** - **28**) to the query structure (**25**) would probably not have been recognized at first sight by a chemist scanning a catalog of available starting materials. In fact, two carbon atoms are marked in Figure 18 to indicate the correspondence of atoms and thus help in the perception of structural similarity.
Can this structural similarity be exploited to develop a synthesis for compound (**25**)? The sequence in Figure 19 shows a potential route for converting the quinoline derivative (**28**) into the compound (**25**) by using standard reactions. This sequence even provides possibilities for controlling the stereochemistry and thus assuring the cis-stereochemistry at the ring junction of rotenone (**19**).

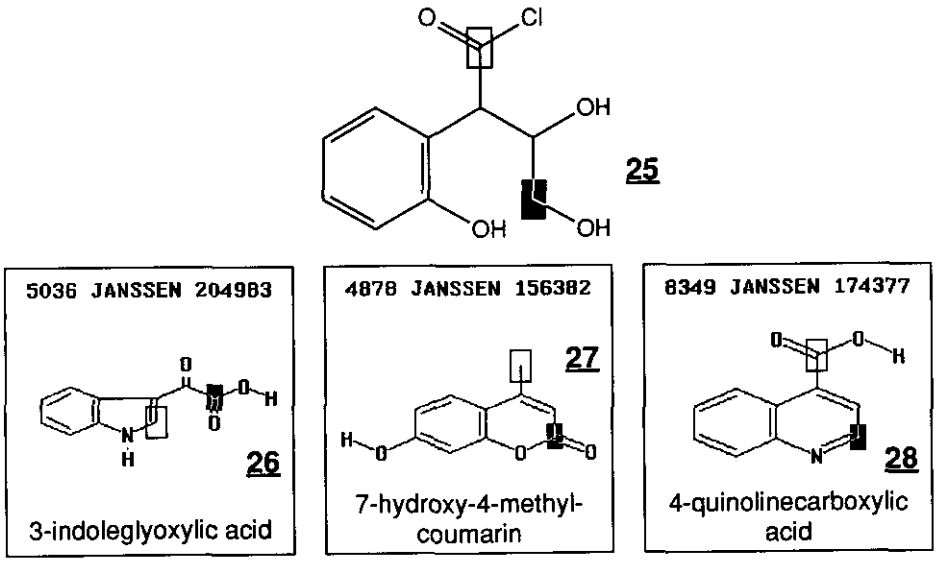

Figure 18 Starting materials having the same carbon skeleton as compound (**25**); one representative of each of the three different heterocyclic ring systems contained in the 45 compounds found is indicated.

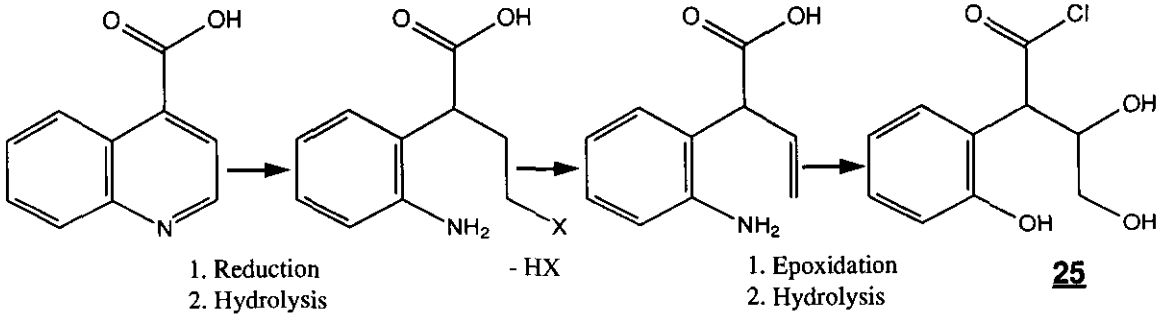

Figure 19 Plausible convertion of quinoline-4-carboxylic acid (**28**) into the synthesis intermediate (**25**).

To summarize, two basic plans for the synthesis of the rotenone skeleton (**19**) have been developed. One finds correspondence in a published synthesis, the other one represents quite a novel approach.

WORKING WITH WODCA

It should have become clear by now that the strength of WODCA lies in its variety of powerful search and planning capabilities: Similarity searching in databases of available starting materials, search and evaluation of strategic bonds, generation of synthesis precursors corresponding to the important synthons, perception and evaluation of stereochemical features, consideration of the price of starting materials etc.

Over and above, WODCA allows the user to rapidly switch between these different search methods to drive the development of a synthesis plan as rapidly as possible to success. In this way, the lateral thinking of a chemist, which is quite different from a rigorously pursued stepwise retrosynthetic analysis is efficiently supported. In order to even enhance this intuitive approach of a chemist the graphical user interface of WODCA has been designed to provide the results of various search methods simultaneously on the screen by making heavy use of the windowing technique.

Figure 21 shows a typical display of the screen during a WODCA study. In the lower right-hand corner of the screen the structure of the synthesis target is displayed. The column of the left-hand side shows icons for various tools that are activated for this synthesis run. Next to it is the synthesis target with indications of strategic bonds. To the right-hand side are structures of available starting materials found by a similarity search. The lower left-hand corner shows the synthesis tree with the target molecule, synthesis precursor, and starting materials, all indicated by little boxes.

Having all these different pieces of information developed during a synthesis search simultaneously available gives the user much insight into the ways to efficiently solve the synthesis problem. Furthermore, by clicking in the appropriate window or opening additional windows she/he can continue searches in many different directions such as the breaking of specific strategic bonds, initiate other similarity searches for starting materials, resume the search at any given point in the synthesis tree, etc. Thus, the user has control of the direction into which a synthesis is developed and can bring in her/his specific needs into a WODCA study.

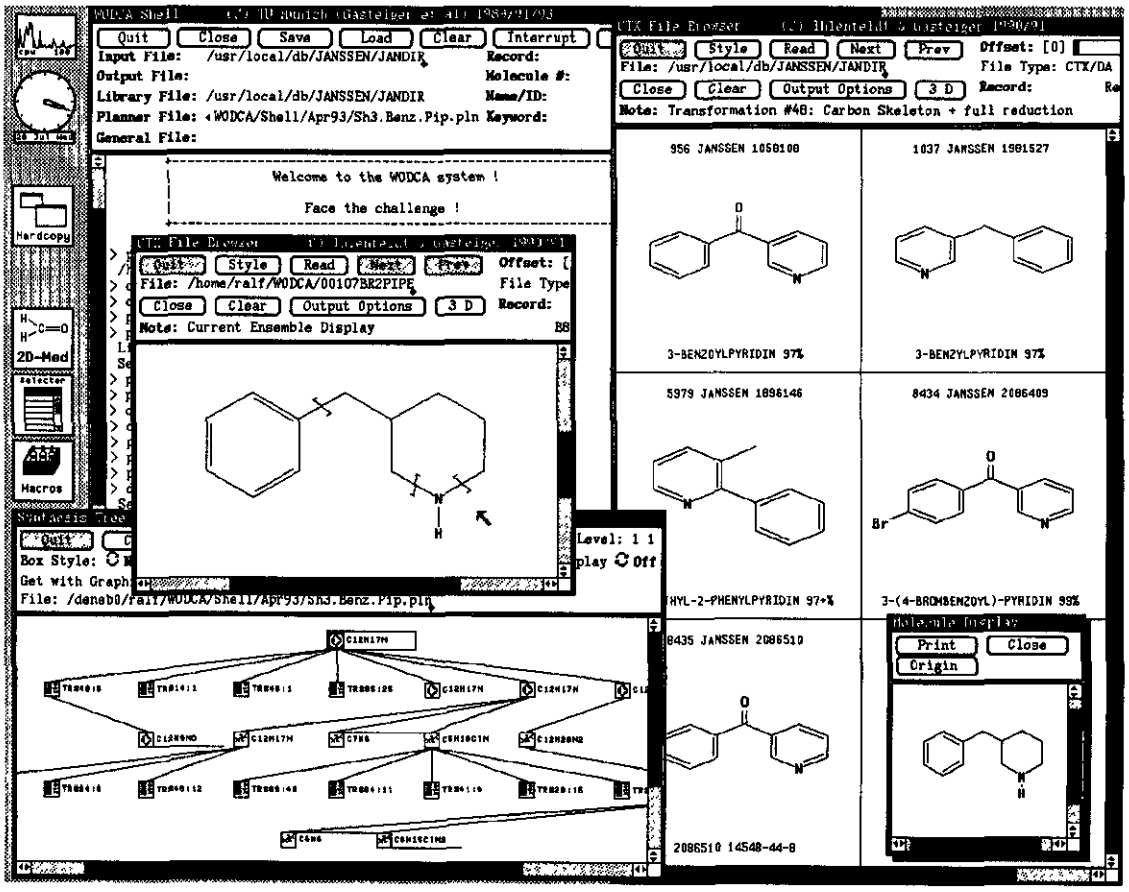

Figure 20 Typical display of the screen during a synthesis study with the WODCA system. Results from the design of a synthesis for 3-benzylpiperidine, discussed in a previous section, are shown.

ACKNOWLEDGMENTS

Our thanks are addressed to Shell Research for supporting the development of WODCA, to Janssen Chimica and to Prof. S. Hanessian for providing us with computer-readable catalogs of available starting materials.

REFERENCES

1. E. J. Corey, *Pure Appl. Chem.*, 1969, **14**, 19.
2. J. Gasteiger, W.-D. Ihlenfeldt, and P. Roese, *Red. Trav. Chim. Pays-Bas*, 1992, **111**, 270.
3. J. Gasteiger, W.-D. Ihlenfeldt, R. Fick, and J. R. Rose, *J. Chem. Inf. Comput. Sci.*, 1992, **32**, 700.
4. W.-D. Ihlenfeldt and J. Gasteiger, *Angew. Chem.*, accepted for publication.
5. S. Hanessian, M. Botta, B. Larouche, and A. Boyaroglu, *J. Chem. Inf. Comput. Sci.*, 1992, **32**, 718.
6. L. A. Walters and S. M. McElvain, *J. Am. Chem. Soc.*, 1933, **55**, 4625.
7. For an overview on rotenone biosynthesis see: L. Crombie, *Nat. Prod. Rep.*, 1984, **1**, 3.
8. S. M. F. Lai, J. J. A. Orchison, and D. A. Whiting, *Tetrahedron*, 1989, **45**, 5895.

Received, 8th August, 1994