STANFORD ARTIFICIAL INTELLIGENCE PROJECT
MEMO AIM-141

COMPUTER SCIENCE DEPARTMENT
REPORT NO. CS-203

THE HEURISTIC DENDRAL PROGRAM FOR EXPLAINING EMPIRICAL DATA

BY

BRUCE G. BUCHANAN

JOSHUA LEDERBERG

CO-PRINCIPAL INVESTIGATORS: E. FEIGENBAUM & J. LEDERBERG

FEBRUARY 1971

COMPUTER SCIENCE DEPARTMENT

STANFORD UNIVERSITY

STANFORD ARTIFICIAL INTELLIGENCE PROJECT
MEMO AIM-141

FEBRUARY 1971

COMPUTER SCIENCE DEPARTMENT REPORT
NO. CS203

THE HEURISTIC DENDRAL PROGRAM FOR EXPLAINING EMPIRICAL DATA*

by

Bruce G. Buchanan
Joshua Lederberg

Co-Principal Investigators: E. Feigenbaum & J. Lederberg

ABSTRACT: The Heuristic DENDRAL program uses an information processing model of scientific reasoning to explain experimental data in organic chemistry. This report summarizes the organization and results of the program for computer scientists. The program is divided into three main parts: planning, structure generation, and evaluation.

The planning phase infers constraints on the search space from the empirical data input to the system. The structure generation phase searches a tree whose termini are models of chemical molecules using pruning heuristics of various kinds. The evaluation phase tests the candidate structures against the original data. Results of the program's analyses of some test data are discussed.

*This research was supported by the Advanced Research Projects Agency (SD-183). Much of the work reported here was performed by Mrs. Georgia Sutherland and Mr. Allan Delfino. The assistance of Dr. Alan Duffield and Professor Carl Djerassi is also gratefully acknowledged.

The Heuristic DENDRAL Program for Explaining Empirical Data

The Heuristic DENDRAL program applies an information processing model of scientific reasoning to a specific problem in organic chemistry. It reasons its way from experimental chemical data to explanatory hypotheses about the molecular structure of compounds. For now, the program ignores other kinds of scientific reasoning such as theory formation: its task is to explain data within an established theory. This report describes the Heuristic DENDRAL program for IFIP members who might have hoped for a succinct description in our artificial intelligence reports (for example, [7],[8], [9]) and who would like to avoid the chemical details found in our publications for chemists [2], [3], [4], [5], [6].

This paper is divided into three main parts: (I) a brief description of the task area, mass spectroscopy; (II) a discussion of the artificial intelligence aspects of the program; and (III) a summary of results.

I. THE TASK AREA

Organic chemists are primarily concerned with either the analysis or synthesis of compounds, that is, with either identifying or manufacturing chemical molecules. Mass spectrometry is a branch of analytic chemistry in which the substance to be identified is vaporized and bombarded with electrons in a mass spectrometer in order to obtain data on the resulting fragmentations.

The data are arranged in a mass spectrum, which shows the masses of fragments produced in the instrument plotted against their relative abundance. Thus the task of the chemist is to use his knowledge of the behavior of molecules in a mass spectrometer to identify the structure of compounds.

The information processing nature of the problem is one important reason for selecting the analysis of mass spectra as the task area. Chemists themselves use non-mathematical models of organic molecules and of the mass spectrometer to analyze mass spectra. They also use many complex judgmental rules. Another reason for selecting a branch of organic chemistry as the program's task area is that a notational algorithm for representing and generating chemical molecules invented by Lederberg [1] is particularly well-suited for computer use. This algorithm, named DENDRAL, is described in section II-B of this paper.

II. PROGRAM ORGANIZATION

Heuristic DENDRAL is organized as a heuristic search program which searches the space of organic molecular structures for the molecule which best explains the, experimental data. The input to the program is the mass spectrum, empirically determined by inserting a sample of a compound into the mass spectrometer. Out of the implicit space of all possible molecular structures the program selects the structures which best explain the data -- often a single structure. Because of the size of the space, it is necessary to reduce the search through the judicious use of heuristics. And, because several structures may be plausible explanations, it is necessary to provide a means for evaluating alternatives.

In test cases, where we know the structure of the sample compound, the program usually produces the correct structure in its answer set. Its pruning

and evaluation heuristics are good enough that this is a small set, as can be seen in the accompanying tables. The working chemist, of course, does not ordinarily know the structure of his sample.

The heart of Heuristic DENDRAL, as of any heuristic search program, is the generator of the search tree. The tree, in this case, is the tree of successive attachments of chemical atoms into larger and larger graph structures. The generator is the DENDRAL algorithm. At the first node of the tree is the initial set of unstructured atoms; deeper levels of the tree correspond to partial structures with more atoms in the structure and fewer unattached atoms. At the ends of all the branches are complete molecular structures with every atom in the initial set allocated to some place in the structure. The DENDRAL algorithm makes all possible attachments of atoms irredundantly at every level, and it provides the capabilities for heuristic pruning of the tree. Constraints on the generator take two forms: search reduction based on plans inferred from the mass spectral data and search reduction based on considerations of chemical stability.

A. Planning: Search Reduction Based on the Mass Spectral Data

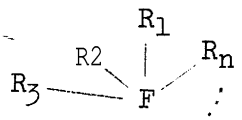
Among the large numbers of molecular structures at the termini of the search tree, planning can describe constraints on the space which are severe enough to limit the number of termini to a few dozen or even just one or two structures. The search reduction power of the plan depends upon the amount of chemical theory embodied in the underlying planning heuristics.

1. Constructing Plans from the Data

A plan is a set of constraints for the generator which limits the output structures to those which are most relevant to the data. The data may be the

mass spectrum or other experimental data on the sample, for example, a nuclear magnetic resonance (NMR) spectrum.

From mass spectral data it is often possible to infer that particular partial structures, or "superatoms" must be contained in each of the candidate structures. And it is often possible to determine the positions of the superatoms within the context of the remaining unstructured atoms. Currently, the program infers the presence of only one superatom at a time, so the form of this part of a plan is



The. F in the center of this scheme is the superatom, which has been identified. (It is called a "functional group" by chemists, thus the "F".) The R's are the weights of the appendant radicals, which surround F . Having this information constrains the search to molecules which conform to the particulars of this scheme.

Plans are constructed by the planning program by means of a complex set of judgmental rules like those used by chemists. Sets of peaks in the mass spectral data often characterize the functional group in the molecule, and thus identify F in the plan. The context of those peaks in the data, then, place the functional group in the molecule relative to the other atoms, and thus identify the R's in the plan. For example, the functional group "ketone" (C=O) can be identified by the existence of a pair of peaks in the spectrum at mass points R_1+28 and R_2+28 whose sum is the molecular weight plus 28 mass units. (A few additional constraints insure that accidental peaks in the data

will not indicate the ketone group. For example, at least one of the peaks must be a prominent peak in the spectrum.) The existence of such a pair of peaks identifies F as a carbon atom doubly bonded to an oxygen atom., The specific values of R_1 and R_2 , say 43 and 71, can then identify the masses of the two radicals appendant from the 'ketone group. Thus the final plan becomes:

$$(43) - \overset{\overset{O}{\|}}{C} - (71) \quad .$$

Other types of data may be employed by the planning program if they are available. For the analysis of amines, for example, data from nuclear magnetic resonance experiments greatly augment the power of the planning program. The tables of results for amines, ethers , alcohols, thioethers and thiols show the dramatic reduction possible when NMR data are used. In these cases the NMR data were used to infer the numbers of methyl (CH_3) radicals present in the test samples and were used to help infer the structures of the superatoms. It will be possible to incorporate judgmental rules to be used with still other kinds of experimental data, as the need arises.

The planning program works best with data from unringed molecules containing a single functional group. The reason for this is that the mass spectrometry theory for these molecules is simpler and less ambiguous than for more complex molecules. The next section digresses somewhat from the present discussion to explain how we have been able to automate the generation of the planning heuristics on the basis of the known theory.

2. Generating Planning Heuristics from the Theory

Some of the most powerful planning heuristics used by chemists (and by the program) were noticed to be relatively straightforward consequences of the theory of mass spectrometry. For the set of molecules containing a single functional group, the planning heuristics can be generated from a few well-known rules of mass spectrometry. We have written a program, external to the Heuristic DENDRAL system, for generating these planning rules.

This external program is in two conceptually distinct parts: a superatom generator and a planning rule generator. The superatom generator is a specialized version of the DENDRAL structure generator mentioned previously. Its task is to construct candidate superatoms for inclusion in the plan. The planning rule generator uses the theory of mass spectrometry to construct a set of heuristics for inferring the presence of each superatom in the mass spectral data. The whole process of constructing plans, then, can be thought of as a problem solving activity where the input is the mass spectrum together with a set of non-carbon atoms that may be in functional groups, and the output is a plan or set of alternative plans for generating candidate structures which explain the data.

3. Summary

Regardless of the source of the candidate superatoms and their planning heuristics, whether from a chemist or from a program, the Heuristic DENDRAL system uses them to construct plans. It tests each candidate functional group (superatom) against the original data by applying the planning heuristics. If the functional group satisfies the criteria, it is put into a plan together with other inferred constraints. The search reduction effect of planning is shown in Tables 2-5.

A severe limitation on this problem solver is that it depends upon knowing that only one superatom containing non-carbon atoms is present in the structure of the sample (ignoring hydrogens) and consequently that only one functional group is present. The theory which the rule generator can use does not always apply when several functional groups are present, nor has much theory been developed to tell the program what does happen. Although chemists consider more complex cases and the generator of superatoms can easily be extended to handle them, the mass spectrometry theory, and consequently the planning rule generator, cannot be so easily extended.

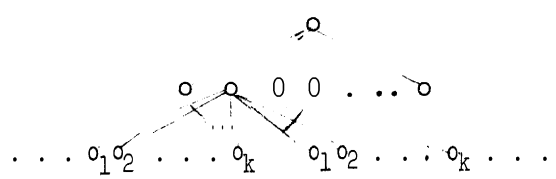
B. Structure Gene-ration

The DENDRAL algorithm provides a representation of objects in the search space -- chemical molecules -- and describes the procedure for generating them. Both the representation and the procedure have proved amenable to computer use, with very few changes. The algorithm uses no other chemical knowledge than the valence, or number of allowable links, for each type of chemical atom. Carbon, for example, has a valence of four, oxygen two, and so forth. Within these mild constraints the algorithm is capable of generating all topologically possible non-ringed graph structures from a given collection of chemical atoms. The actual canons of procedure will not be discussed here. The important point to note is that the algorithm's output of topologically possible molecular structures can contain a very great number of structures which are implausible from the standpoint of chemical stability. Search reduction heuristics on the list known as BADLIST prune the tree as unstable chemical structures begin to emerge. This reduction can be seen from Table 1. In the other cases BADLIST has no effect unless a chemist wishes to change it so as to prune some structures

which are now allowed.

1. The Search Space

The search space itself is organized as an AND/OR tree which is searched depth-first. The first level of the tree, after the specification of the initial collection of atoms, is the set of all possible molecule centers, or centroids. Because any one of these centroids may lead to the solution of the program, this level is a set of OR nodes. Also, for this reason, the OR nodes are ordered by the program so that the most likely centroid appears first in the set. The. next level of the tree, just beyond the node specifying a possible centroid, specifies the possible ways the remaining atoms in the composition can be partitioned to the unfilled links of the centroid. A central carbon atom with three unfilled links, for example, must be completed by having three radicals, made from the remaining composition, attached to the links. Thus, beyond that node the program will grow several sets of AND nodes, each set defining a possible partition of the remaining atoms into three clusters. The scheme of the tree generation for these two levels is shown in the diagram below.



For each AND set of subproblems, all of which must be successfully completed if the program is to grow the tree beyond any of the nodes, the program attempts the most difficult subproblem first. That is, it orders the clusters of atoms

from the superatom as centroid, the program explores only that part of the tree in which the primary partitions of the remaining atoms are compatible with the radical weights specified in the plan. Consider again the planning example considered in part (A), where the plan was

$$(43) - \overset{O}{\overset{\|}{C}} - (71) \quad .$$

This means that generation proceeds by first removing a carbon and an oxygen atom from the initial set of atoms and then constructing only the partitions of the remaining atoms which are compatible with weights 43 and 71, that is, partitions of C_3H_7 and C_6H_{11} .

Although rarely used, the ability to accept a chemist's intuitions or biases is a powerful search reduction tool. BADLIST itself reflects one scientist's intuitions about the subgraphs responsible for unstable structures. But beyond that, it is easy for an individual to guide the search by adding (or deleting) constraints to RADLIST and GOODLIST. A chemist can supress all occurrences of a superatom from the generator's output by adding that superatom to BADLIST. Conversely, he can force the occurrence of a superatom in every output structure by adding that superatom to GOODLIST.

The Structure Generator is the central part of the total Heuristic DENDRAL program. It was mentioned earlier that the planning program can often specify such a detailed plan that only a single structure fits the plan. In spite of this power it is necessary to retain the capabilities of a general heuristic search program to deal with cases outside the scope of the Planner's power. The output of the Structure Generator is a list of molecular structures. They are all plausible candidates for explaining the given mass spectrum because

they are all chemically stable and they all fit the constraints of the plan inferred from the experimental data.

c. Evaluation

The purpose of the last phase of the Heuristic DENDRAL program is to cull the least promising of the plausible candidate structures and rank the remaining ones. For both of these jobs the program obtains a predicted mass spectrum from its internal model of the mass spectrometer. The significant peaks in the predicted mass spectrum are then matched against the original spectrum. A candidate structure is rejected-if its predicted spectrum is inconsistent with the original data, and the remaining candidates are ranked by how well they explain the original data.

The prediction program, known as the Predictor, consists of two main parts: a theory of mass spectrometry plus a large number of routines for describing mass spectrometric processes and manipulating molecular structures in accordance with those processes. It is not necessary to describe the details of either of these parts, but the separation of theory from the rest of the program is of some interest.

By separating the theory of mass spectrometry in the Predictor from the routines which reference it, the theory is much easier to change -- either by hand or by a program. The theory is a set of data which the program sorts through to determine the actions to perform and the parameter settings associated with those actions. The theory embodied in this data structure is organized as situation-action rules (or productions). The program checks for the truth of each situation in the current context and, if true, executes the associated set of actions. For example, the Predictor checks for the occurrence of the

ketone functional group by looking for the subgraph C=O in the graph structure of the molecular structure. If the subgraph is present the program executes routines for performing cleavages and rearrangement processes characteristic of 'ketones.

The input to this last phase of the program is a set of molecular structures; the intermediate result is a set of predicted mass spectra, and the final output is a ranked list of structures which are consistent with the original data. Consistency, in this case, means that every significant feature of the predicted spectrum for the candidate structure actually appears in the original data. Thus the predictive test can only disconfirm candidates. Scoring the candidates on the basis of how many peaks in the original data they can explain is meant to estimate degrees of confirmation. The score for a candidate is the sum of the significance weightings assigned to its predicted mass spectrum by the program. Thus a candidate which explains peaks thought to be very significant will rank higher than one which explains as many (or possibly more) peaks of less significance.

III. RESULTS

Although results of the program's analyses of selected mass spectra have been published in chemistry journals (see [2] -[6]) they have not been adequately summarized for computer scientists. The accompanying tables show the sizes of the problem spaces for different classes of problems and the search reduction achieved by the program.

The amino acids shown in Table 1 were analyzed without planning, but with references to the data during structure generation by a simple theory called

the "zero-order" theory. Amino acids are characterized by the presence of both a nitrogen and a carboxylic acid group (-C=O) in the molecule. They happen
'OH..

to lend themselves to this simple kind of analysis because they tend to'fragment in almost every possible way in a mass spectrometer, just as the zero-order theory predicts. This is not true of other classes of compounds. BADLIST is able to constrain the size of the search space dramatically, as noted by the difference between the columns entitled "Number of Possible Structures" and "Number of Plausible Structures", because more than one non-carbon atom is present in amino acids.=. This desirable reduction is lost in the other cases, as indicated in the footnote to the third column of Tables 2-5.

For the ketones, shown in Table 2, planning was necessary to achieve the search reduction noted between the columns entitled "Number of Plausible Structures" and "Number of Structures Generated". Applying a few well-known rules of mass spectrometry was almost solely responsible for this reduction. Other rules about the mass spectrometric behavior of ketones allowed the evaluation program to exclude some of the candidates generated and successfully rank the-remaining ones. As noted before, ketones are characterized by the presence of the chemical substructure c=o .

Tables 3-5 show the results of the program's analysis of unringed compounds containing the substructures

-N-	(amines)
-O-	(ethers)
-OH	(alcohols)
-S-	(thioethers)
-SH	(thiols) .

For all of this work the planning program contained a much larger body of theoretical knowledge than in the ketone case. Its theory about the mass spectrometry of these classes of compounds, in fact, was as complete as the theory in the Predictor. And it included nuclear magnetic resonance (NMR) theory which the Predictor does not. Thus, the plans which it was able to construct were so detailed that the evaluation phase could make no further improvements. In other words, there was no theory left to use for evaluation which had not already been used in planning.

IV. CONCLUSION

The Heuristic DENDRAL program successfully explains experimental data for many test problems in analytic organic chemistry. On a limited class of molecules it performs at about the same level as a post-doctoral chemist. However, the class of problems which can be solved is still very small relative to those a practicing chemist may see. Much of our future work will be devoted to extending the power of the program to cover, for example, compounds with several functional groups and compounds containing an arbitrary number of rings. We anticipate much work, also, on extending the program to cover more varied kinds of scientific reasoning.

TABLE 1.

Amino Acid Results - without prior Planning

Name of "Unknown" Amino Acid	Chemical Formula	Number of Possible Structures (1)	Number of Plausible Structures (2)	Number of Structures Generated (3)	Rank Order of Correct Candidate (4)
Glycine	C2H5NO2	38	12	8	1st, 7 excluded
Alanine	C3H7NO2	216	50	3	1st
Ser ine	C3H7NO3	324	40	10	1st, 9 excluded
Threonine	C4H9NO3	1758	238	2	1st
Leucine	C6H13NO2	10000 (approx.)	3275	288	Tied for 2nd, 277 excluded

(1) The total number of possible structures is the number of topologically possible (and distinctive) molecular structures generated by the algorithm within valence considerations alone.

(2) The number of plausible structures is the number of molecular structures in the total space which also meet the a priori conditions of chemical stability on BADLIST. The a priori rules have greater effect with increased numbers of non-carbon, non-hydrogen atoms,

(3) The number of structures generated is the number of molecular structures actually generated by the program as candidate explanations of the experimental data. Pruning has been achieved by using the "zero-order" theory during structure generation.

(4) The rank order of the correct structure is the evaluation program's assignment of rank to the actual molecular structure used as a test "unknown". The number of structures excluded in the validation process is also indicated.

TABLE 2

Ketone Results - with Prior Planning and Post-Evaluation

Name of "Unknown" Ketone	Chemical Formula	Number of Plausible Structures (1)	Number of Structures Generated (2)	Rank Order of Correct Candidate (3)
2-Butanone	C4H8O	11	1	1st
3-Pentanone	C5H10O	33	1	1st
3-Hexanone	C6H12O	91	1	1st
2-Methyl-hexan-3-one	C7H14O	254	1	1st
3-Heptanone	C7H14O	254	2	Tied for 1st
3-Octanone	C8H16O	698	4	1st
4-Octanone	C8HI6O	698	2	1st, 1, excluded
2,4-Dimethyl-hexan-3-one	C8H16O	69%	4	Tied for 1st, 1 excluded
6-Methyl-heptan-3-one	C8H16O	698	4	1st
3-Nonanone	C9H18O	1936	7	1st
?-Methyl-octan-3-one	C9H18O	1936	4	1st (4)
4-Nonanone	C9H18O	1936	4	1st (4)

(1) The number of plausible structures is the number of molecular structures in the total space which also meet the a priori conditions of chemical stability on BADLIST. The a priori rules have no effect with formulas containing a single non-carbon, non-hydrogen atom. Thus, this column also represents the total number of possible structures.

(2) The number of structures generated is the number of molecular structures actually generated by the program as candidate explanations of the experimental data. Pruning has been achieved by using the planning. information from the Planning Program.

(3) The rank order of the correct structure is the evaluation program's assignment of rank to the actual molecular structure used as a test "unknown". The number of structures excluded in the process is also indicated.

(4) Previous publication showed the correct structure excluded. The general rules of the program have since been modified to improve its performance.

TABLE 3

Amine Results - with Prior Planning but No Post-Evaluation

Name of "Unknown" Amine	Size: Cn	Number of Plausible Structures (1)	Number of Structures Generated (2) MS	NMR	Name of "Unknown" Amine	Size: Cn	Number of Plausible Structures (1)	Number of Structures Generated (2) MS	NMR
n-propyl	C3	4	2	1	N-methyl-di-iso-propyl	C7	89	15	3
iso-propyl		4	2	1	n-octyl	C8	211	39	1
n-butyl	c4	8		1	Ethyl-n-hexyl		211	24	1
iso-butyl		8	2	1	l-methyl heptyl		211	34	1
Sec-butyl		8	4	2	2-ethyl hexyl		211	39	9
Tert-butyl		8	3	1	1,1-dimethyl hexyl		211	32	4
Di -ethyl		8	3	1	Di-n-butyl		211	24	1
N-methyl-n-propyl		8	4	1	Di-sec-butyl		211	33	8
Ethyl-n-propyl	C5	17	5	1	Di-iso-butyl		211	17	5
:l-methyl -di-ethyl		17	4	1	Di -ethyl-n-butyl		211	17	3
n-pentyl		17	4	1	3-octyl		211	26	2
iso-pentyl		17		2	n-nony 1	C9	507	89	1
2-pentyl		17	2	1	Y-methyl -di -n-butyl		507	13	1
3-pentyl		17	4	1	Tr i -n-propyl		507	2	1
3-methyl-2-butyl		17	4	1	Di-n-pentyl	C10	1238	83	1
N-methyl -n-butyl		17		1	Di -i so-pentyl		1238	109	16
N-methyl -sec-butyl		17	3	1	N,N-dimethyl-2-ethyl hexyl		1238	156	9
N-methyl-iso-butyl		17	4	1	n-undccyl	C11	3057	507	1
n-hexy 1	C6	39	8	1	n-dodecyl	C12	7639	1238	1
Tr i-ethyl		3 9	8	1	n-tetradecyl	C14	48865	10115	1
2-hexyl		39	8	1	Di -n-heptyl		48865	646	1
Di-n-propyl		39			N,N-dimethyl-n-dodecyl		48865	4952	1
Di-iso-propyl		39	8	1	Tri-n-pentyl	C15	124906	40	1
N-methyl -n-pentyl		39	8	11	Bi s-2-ethyl hexyl	C16	321988	2340	24
N-methyl-iso-pentyl		39	8	2	N,N-dimethyl -n-tetradecyl		321988	3895	1
Ethyl-n-butyl		39	6	1	Di-ethyl-n-dodecyl		321988	2476	1
N,N-dimethyl-n-butyl		59	10	1	n-heptadecyl	C17	830219	124906	1
n-heptyl	C7	89	17	1	N-methyl-bis-2-ethylhexyl		830219	2340	24
Ethyl -n-pentyl		89	16	1	n-octadecyl	C18	2156010	48865	1
n-butyl-iso-propyl		89	11	1	N-methyl -n-octyl-n-nonyl		2156010	15978	1
4-methyl-2-hexyl		89	16	4	N,N-dimethyl-n-octadecyl	C20	14715813	1284792	1

(1) The number of plausible structures is the number of molecular structures in the total space which also meet the a priori conditions of chemical stability on BADLIST. The a priori rules have no effect with formulas containing a single non-carbon, non-hydrogen atom. Thus, this column also represents the total number of possible structures.

(2) The number of structures generated is the number of molecular structures actually generated by the program as candidate explanations of the experimental data. Pruning has been achieved by using the planning information from the Planning program.

MS = Number of structures when only mass spectrometry is used in planning.
NMR = Number of structures when NMR data are used in planning to infer the number of methyl radicals.

TABLE 4

Ether and Alcohol Results - with Prior Planning but No Post-Evaluation

Name of "Unknown" Alcohol	Size: Cn	Number of Plausible Structures (1)	Number of Structures Generated (2) MS	Number of Structures Generated (2) NMH
n-butyl	c4	7	2	1
Iso-butyl			2	1
Sec-butyl	C5	14 3	3	2
2-methyl-2-butyl			1	1
n-pentyl		14	4	1
3-pentyl		14	1	1
2-methyl-1-butyl		14	4	2
2-pentyl		14	2	1
3-hexyl	C6	32	2	1
3-methyl-1-pentyl		32	8	4
4-methyl-2-pentyl		32	4	1
n-hexyl		32	8	1
3-heptyl	c7	72	4	1
2-heptyl		72	8	1
3-ethyl-3-pentyl		72	1	1
2,4-dimethyl-3-pentyl		72	3	1
n-heptyl		72	17	1
3-methyl-1-hexyl		72	17	6
n-octyl	C8	171	39	1
3-octyl		171	8	1
2,3,4-trimethyl-3-pentyl		171	3	1
n-nonyl	C9	405	89	1
2-nonyl		405	39	1
n-decyl	C10	989	211	1
6-ethyl-3-octyl		989	39	9
3,7-dimethyl-1-octyl		989	211	41
n-dodecyl	C12	G045	1238	1
2-butyl-1-octyl		6045	1238	25
n-tetradecyl	C14	38322	7639	1
3-tetradecyl		38322	1238	1
n-hexadecyl	C16	151375	4 8 8 6'5	1

Name of "Unknown" Ether	Size:,	Number of Plausible Structures (1)	Number of Structures Generated (2) MS	Number of Structures Generated (2) NMR
Methyl-n-propyl	c4	7	2	1
Methyl-iso-propyl		7	3	1
Methyl-n-butyl		14	2	1
Methyl-iso-butyl		14	2	1
Ethyl-iso-propyl		14	1	1
Ethyl-n-butyl	C6	32	4	1
Ethyl-iso-butyl		32	4	2
Ethyl-sec-butyl		32	2	2
Ethyl-tert-butyl		32	1	1
Di-n-propyl		32	1	1
Di-iso-propyl		32	1	1
n-propyl -n-butyl	c7	72	2	1
Ethyl-n-pentyl		72	4	1
Methyl-n-hexyl		72	8	1
Iso-propyl-sec-butyl		72	3	2
Iso-propyl-n-pentyl		171	4	1
n-propyl -n-pentyl		171	4	1
Di-n-butyl		171	3	1
Iso-butyl-tert-butyl		171		1
Ethyl -n-heptyl	C9	405	32	1
n-butyl -n-pentyl		405	8	1
Di-n-pentyl	C10	989	10	1
Di-iso-pentyl		989	18	7
Di-n-hexyl	C12	6045	125	2
Di-n-octyl	C16	151375	780	1
Bis-2-ethylhexyl		151375	780	21
Di-n-decyl	C20	11428365	22366	1

(1) The nunber of plausible structures is the number of molecular structures in the total space which also meet the a priori conditions of chemical stability on BADLIST. The a priori rules have no effect with formulas containing a single non-carbon, non-hydrogen atom. Thus, this column also represents the total number of possible structures.

(2) The number of structures generated is the number of molecular structures actually generate? by the program as candidate explanations of the experimental data. Pruning has been achieved by using the planning information from the Planning program.

MS = Number of structures when only mass spectrometry is used in planning.
NMR = Number of structures when NMR data are used in, planning to infer the number of methyl radicals.

TABLE 5

fhioether and Thiol Results - with Prior Planning but No Post-Evaluation

Name of "Unknown" Thioether	Size: Cn	Number of Plausible Structures (1)	Number of Structures Generated (2) MS	NMR	Name of "Unknown" Thiol	Size: Cn	Number of Plausible Structures (1)	Number of Structures Generated (2) MS	NMR
Methyl-ethyl	c3	3	1	1	n-propyl	c3	3	2	1
Methyl-n-propyl	C4	7	1	1	iso-propyl		3	1	1
Methyl -iso-propyl		7	2	1	n-butyl	c4	7	3	1
Di-ethyl		7	1	1	iso-butyl		7	3	1
Methyl-n-butyl	C5	14	3	1	Tert-butyl		7	1	1
Methyl-iso-butyl		14	5	2	2-methyl-2-butyl	c5	14	1	1
Methyl -tert-butyl		14	1	1	3-methyl-2-butyl		14	2	1
Ethyl-iso-propyl		14	1	1	3-methyl-1-butyl		14	6	3
Ethyl-n-propyl		14	2	1	n-pentyl		14	4	1
Ethyl-n-butyl	C6	32	3	1	3-pentyl		14	5	3
Ethyl-tert-butyl		32	1	1	2-pentyl		14	6	3
Ethyl-iso-butyl		32	3	2	n-hexyl	C6	32	8	1
Di-n-propyl		32	2	1	2-hexyl		32	12	5
Methyl-n-pentyl		32	10	1	2-methyl-1-pentyl		32	8	4
Di-iso-propyl		32	1	1	4-methyl-2-pentyl		32	4	2
Ethyl-n-pentyi	C7	72	4	1	3-methyl -3-pentyl		32	1	1
n-propyl-n-butyl		72	5	1	2-methyl-2-hexyl	c 7	72	8	3
iso-propyl -n-butyl		72	5	2	n-heptyl		72	17	1
iso-propyl-tert-butyl		2	1	1	2-ethyl-1-hexyl	C8	171	39	9
n-propyl-iso-butyl		72	3	2	n-octyl		171	39	1
iso-propyl-sec-butyl	7	2	4	3	1-nonyl	C9	405	89	1
n-propyl-n-pentyl	C8	171	4	1	n-decyl	C10	989	211	1
Ethyl-n-hexyl		171	8	1	n-dodecyl	CI2	6045	1238	1
Di-n-butyl		171	5	1					
Di-sec-butyl		171	3	1					
Di-iso-butyl		171	3	1					
Methyl-n-heptyl		171	21	1					
Di-n-pentyl	C10	989	12	1					
Di-n-hexyl	CI2	6045	36	1					
Di-n-heptyl	CI4	38322	153	1					

(1) The number of plausible structures is the number of molecular structures in the total space which also meet the a priori conditions of chemical stability on BADLIST. The a priori rules have no effect with formulas containing a single non-carbon, non-hydrogen atom. Thus, this column also represents the total number of possible structures.

(2) The number of structures generated is the number of molecular structures actually generated by the program as candidate explanations of the experimental data. Pruning has been achieved by using the planning information from the Planning program.

MS = Number of structures when only mass spectrometry is used in planning.
NMR = Number of structures when NMR data are used in planning to infer the number of methyl radicals.

BIBLIOGRAPHY

1. J. Lederberg and E.A. Feigenbaum, "Mechanization of Inductive Inference in Organic Chemistry". In Formal Representation of Human Judgment (B. Kleinmuntz, ed.), John Wiley & Sons, Inc., 1968. (Also Stanford' Artificial Intelligence Project Memo No. 54.)

2. J. Lederberg, G.L. Sutherland, B.G. Buchanan, E.A. Feigenbaum, A.V. Robertson, A.M. Duffield, and C. Djerassi, "Applications of Artificial Intelligence for Chemical Inference I. The Number of Possible Organic Compounds: Acyclic Structures Containing C, H, O and N". Journal of the American Chemical Society, 91,2973-2976(1969).

3. A.M. Duffield, A.V. Robertson, C. Djerassi, B.G. Buchanan, G.L. Sutherland, E.A. Feigenbaum, and J. Lederberg, "Applications of Artificial Intelligence for Chemical Inference II. Interpretation of Low Resolution Mass Spectra of Ketones". Journal of the American Chemical Society, 91,2977-2981(1969).

4. G. Schroll, A.M. Duffield, C. Djerassi, B.G. Buchanan, G.L. Sutherland, E.A. Feigenbaum, and J. Lederberg, "Application of Artificial Intelligence for Chemical Inference III. Aliphatic Ethers Diagnosed by Their Low Resolution Mass Spectra and NMR Data". Journal of the American Chemical Society, 91, 7440-7445(1969).

5. A. Buchs, A.M. Duffield, G. Schroll, C. Djerassi, A.B. Delfino, B.G. Buchanan, G.L. Sutherland, E.A. Feigenbaum, and J. Lederberg, "Applications of Artificial Intelligence for Chemical Inference IV. Saturated Amines Diagnosed by Their Low Resolution Mass Spectra and Nuclear Magnetic Resonance Spectra". Journal of the American Chemical Society, 92,6831 (1970).

6. A. Buchs, A.B. Delfino, A.M. Duffield, C. Djerassi, B.G. Buchanan, E.A. Feigenbaum, and J. Lederberg, "Applications of Artificial Intelligence for Chemical Inference VI. Approach to a General Method of Interpreting Low Resolution Mass Spectra with a Computer". Chemica Acta Helvetica, 53,1394(1970).

7. B.G. Buchanan, G.L. Sutherland, and E.A. Feigenbaum, "Heuristic DENDRAL: A Program for Generating Explanatory Hypotheses in Organic Chemistry". In Machine Intelligence 4 (B. Meltzer and D. Michie, eds.) Edinburgh University Press (1969). (Also Stanford Artificial Intelligence Project Memo No. 62.)

8. B.G. Buchanan, G.L. Sutherland, and E.A. Feigenbaum, "Toward an Understanding of Information Processes of Scientific Inference in the Context of Organic Chemistry". In Machine Intelligence 5, (B. Meltzer and D. Michie, eds.) Edinburgh University Press (1969). (Also Stanford Artificial Intelligence Project Memo No. 99.)

9. E.A. Feigenbaum, B.G. Buchanan, and J. Lederberg, "On Generality and Problem Solving: A Case Study Using the DENDRAL Program". In Machine Intelligence 6 (B. Meltzer and D. Michie, eds..) Edinburgh University Press (in press). (Also Stanford Artificial Intelligence Project Memo No. 131.)