

Heuristic Theory Formation: Data Interpretation and Rule Formation

B. G. Buchanan, E. A. Feigenbaum and N. S. Sridharan

Computer Science Department
Stanford University

I. INTRODUCTION

Describing scientific theory formation as an information-processing problem suggests breaking the problem into subproblems and searching solution spaces for plausible items in the theory. A computer program called meta-DENDRAL embodies this approach to the theory formation problem within a specific area of science.

Scientific theories are judged partly on how well they explain the observed data, how general their rules are, and how well they are able to predict new events. The meta-DENDRAL program attempts to use these criteria, and more, as guides to formulating acceptable theories. The problem for the program is to discover conditional rules of the form $S \rightarrow A$, where the S 's are descriptions of situations and the A 's are descriptions of actions. The rule is interpreted simply as 'When the situation S occurs, action A occurs'.

The theory formation program first generates plausible A 's for theory sentences, then for each A it generates plausible S 's. At the end it must integrate the candidate rules with each other and with existing theory. In this paper we are concerned only with the first two tasks: data interpretation (generating plausible A 's) and rule formation (generating plausible S 's for each A).

This paper describes the space of actions (A 's), the space of situations (S 's) and the criteria of plausibility for both. This requires mentioning some details of the chemical task since the generators and the plausibility criteria gain their effectiveness from knowledge of the task.

The theory formation task

As in the past, we prefer to develop our ideas in the context of a specific task area. Thus the computer program under development works with data from organic mass spectrometry, a subset of analytic chemistry, to infer rules for

the theory of mass spectrometry. The data, from which the program will form rules, are a collection of structural descriptions of chemical molecules – each paired with the analytic data produced by the molecule in an instrument known as a mass spectrometer. The analytic data are usually presented in a fragment-mass table (FMT), or mass spectrum – a table of masses of molecular fragments and their relative abundances.

The program is given some knowledge of the task area. In particular, it is given the ability to manipulate topological descriptions of molecules; it is given primitive terms from which descriptions of causal explanations (of the form $S \rightarrow A$) can be written. But it is not given the predictive rules describing the operation of the mass spectrometer: these are essentially the rules it must discover.

As the discussion develops, we hope to make explicit exactly what knowledge is given to the program. We strongly believe that the performance of artificial intelligence programs increases as the knowledge available to them increases. So we have deliberately given it knowledge of the task area. On the other hand, giving it too much knowledge removes the difficulty (and the interest) from the theory formation problem.

Special purpose program

In order to try out some of our ideas on theory formation quickly, we decided to implement a rather special-purpose program before writing the general theory formation program discussed previously (Buchanan, Feigenbaum and Lederberg 1971). This strategy has three main advantages.

(1) Communication between an expert and a novice seems to be more fruitful the more specific the problem under discussion. At least for purposes of discovering the expert's heuristics and asking him to judge the acceptability of results, this is so.

(2) Putting more knowledge of the task area into the program avoids many of the initial clumsy errors of general-purpose programs. This allows us to work on some of the main problems of theory formation before having to construct an elaborate program substructure.

(3) Working on a specific problem also allows us to produce intermediate results of real value to the practising scientist.

Specifically, we have limited our attention to the class of molecules known as estrogenic steroids, well known for their use in oral contraceptives. The molecules from which rules are constructed are all assumed to be of the same class, with a common skeleton shown in figure 1. Instances of this class have additional atoms (substituents) bonded to the skeleton, replacing hydrogen atoms. One instance is shown in figure 2. Although we believe the program is not limited to just this class, we have no way of knowing this until we attempt rule formation for other classes. See pp. 275–80.

2. CONCEPTUAL FRAMEWORK

The program searches a space of theory sentences (or rules) to construct a collection of sentences explaining the experimental data. The nature of this space is described briefly in this section. Appendixes 1 and 2 define the concepts mentioned here more precisely.

Space of theories

We count as a theory a set of predictive rules in conditional form: $S \rightarrow A$. Although a mere collection of rules often does not warrant the laudatory term 'theory', for purposes of defining the task this simple definition will serve. Exploring the space of theories, then, is exploring the space of rule sets where each rule is of the $S \rightarrow A$ form. The S 's and A 's, in turn, are members of their own separate spaces, so the space of rules is the cross product of the space of S 's and the space of A 's.

As mentioned above, plausible and interesting actions (A 's) are generated first. This is described in detail in the section on data interpretation. Then for each A , plausible and interesting situations (S 's) are generated to form alternative rules of the $S \rightarrow A$ form. The program keeps those rules which explain the sample data most simply and completely, and which still allow extrapolation to unobserved data.

The predictive nature of the theory partially dictates the form, as do the other restrictions we initially place on the kinds of theories we want the program to formulate. Appendix 1 characterizes the space of theories more completely.

We are focusing our attention on theories that will help us predict FMT's (mass spectra) for molecules given their chemical structure. In Appendix 1 we begin by defining what is a predictive theory of mass spectrometry (Appendix 1, Part 1). We indicate that the theory is constructed in several stages. First, the data points in the FMT for each molecule are explained as fully as possible. The general rules formulated must be consistent with all of these individual explanations. We restrict the form of these individual explanations by restricting the terms from which explanatory actions can be described. This leads to the idea of an action-based theory (Appendix 1, Part 2). Also for simplicity, the program is restricted to predicting occurrence and non-occurrence of fragment-mass points, without estimating intensities (y -coordinates) for those data points. This kind of theory we call a 0, 1 theory (Appendix 1, Part 3). The samples we study represent not the entire space of molecules, but rather a restricted class of molecules and thus we introduce the concept of partial theories (Appendix 1, Part 4). The range of validity and applicability of a partial theory is only the space formed by a restricted class of molecules. For each molecule in its range a partial theory may leave some of the peaks in the data unexplained.

Space of actions

The search space is generated from the two primitive 'legal moves': bond cleavage and group migration. Plausible actions, or 'moves', are generated by imposing heuristic constraints on the legal move generator. It should be noted that the plausible actions are generated only once for the basic skeleton common to all molecules. For each molecule, the program records the evidence (or lack of it) for each of the actions. A few precise definitions will help at this point:

- (a) Action = Cleavage followed by Fragment Selection followed by Migration or a set of cleavages followed by fragment selection followed by migration.
- (b) Cleavage = The process of removing all the bonds in a cleavage set.
- (c) Cleavage Set = A set of bonds in the molecule or fragment (edges of a graph) which are necessary and jointly sufficient for separating the molecule into two distinct parts. Bonds to hydrogen atoms are excluded.
- (d) Fragment Selection = the identification of one of the parts of a graph resulting from cleavage.
- (e) Migration = The process of moving a specified group of atoms from one part of the molecule or fragment to another. (A null group is allowed.)

The generation of actions can be simply described now by the following steps.

1. Generate all plausible cleavages (that is, all plausible cleavage sets). (See table 1.)
2. Select each of the two fragments.
3. Specify all plausible actions containing one cleavage followed by migration.
4. Specify all plausible actions containing two cleavages followed by migration.
5. Specify all plausible actions with n cleavages followed by migration. (See table 3.)

The constraints on the generator which determine plausibility of cleavages and of actions are of three kinds, listed below. Table 3 shows the plausible ways of putting actions together from the possible cleavage sets (followed by fragment selection and migration) shown in table 1.

1. *Topological Constraint*

For bond cleavage, consider only sets of bonds which separate the molecular graph into two pieces. These are called cleavage sets.

2. *Chemical Constraints*

- (a) Do not break a set of bonds together if
 - (i) any bonds are in an aromatic ring,
 - (ii) two of the bonds are attached to the same carbon atom,
 - (iii) any bonds (except ring junctures) are other than single bonds.
- (b) Select fragments with k or more carbon atoms only. (Currently $k=6$.)
- (c) Do not allow an additional cleavage in an n -ary action if:
 - (i) the $n+1$ cleavage sets are not disjoint (some bonds are duplicated);

(ii) there is not sufficient evidence for all $n+1$ cleavages (where 'sufficient evidence' may be defined in various ways). For table 1 this heuristic was not used.

(d) Transfer from -2 to $+2$ hydrogens (only) after a set of cleavages (but not in the middle of a set); that is, limit migration to hydrogen atoms, and allow only -2 , -1 , 0 , 1 , or 2 atoms to migrate. (Negative integers specify numbers of atoms lost from the fragment, positive integers specify numbers of atoms gained. Zero indicates no migration, or migrating a null group, if one prefers.)

3. Methodological Constraints

(a) If two sets of actions explain the same data point, ignore the more complex set (Occam's Razor).

(b) Confine cleavages to bonds within the skeleton.

Space of situations

A single situation is a description of a class of molecules, so the space of all possible S 's is the space of all class descriptions. Some of the terms used in defining the space are explained below.

A *class description* is a Boolean combination of features.

A *feature* is a substituent label followed by a position number. This definition of 'feature' holds only for the special purpose program which assumes a common skeleton with numbered positions for placing additional groups of atoms. In general, features can be described in terms of sub-graphs.

A *substituent label* is an arbitrary label describing an atom or group of atoms (substituent) attached to the structural skeleton.

A *position number* is one of the numbers assigned to the nodes of the structural skeleton.

In principle, the generator of situations is quite simple. It must be severely constrained, however, in order to limit the number of S 's actually considered. Two main kinds of constraints are considered: constraints on features and constraints on combinations of features.

1. Constraint on Features

Consider only features which appear in the sample; that is, consider features which mention only substituent labels and position numbers appearing together in some molecule of the sample.

2. Constraints on Boolean Combinations of Features

(a) Allow union and intersection operators only.

(b) Allow only one occurrence of each position number in a combination of features.

(c) Allow only combinations which cover the molecules in the sample; that is, every positive instance in the sample is an instance of the class defined by the combination, or every negative instance is.

(d) Allow only combinations which are consistent with the data; that is,

the class defining positive instances does not also cover known negative instances, and vice versa.

(e) Generate only parsimonious combinations of features; that is, limit the number of terms and the number of Boolean operators.

3. DETAILS OF THE DATA INTERPRETATION PROGRAM

Data interpretation is an essential aspect of theory formation for at least two reasons:

(1) At this early stage, a large volume of experimental data can be compressed into readable results tables. In mass spectrometry, data reduction takes place in many stages. The so-called 'raw data' are initially represented as a continuous graph of detector voltage plotted against time. A sequence of numerical algorithms reduce the data to:

(a) a digital (bar) graph of integrated detector voltages plotted against time of recording the peak centers.

(b) a digital (bar) graph of integrated detector voltages plotted against atomic mass units of the peak centers.

(c) a digital plot of integrated detector voltages *vs.* elemental compositions of fragments (each elemental composition derived from exact mass). Step (c) can be applied to so-called 'high resolution' data, that is, data with precise resolution of the masses accurate to 3 or 4 decimal places.

(d) (c) with detector voltages normalized.

(e) (d) with isotopic contributions deleted.

All of these reductions are made for the routine analysis of mass spectra using the existing corpus of knowledge. Therefore, it is reasonable to start with data already reduced in these ways when attempting to extend the corpus of knowledge. This is especially so because extensions to the corpus will be made using only the primitive concepts which are currently used.

(2) Here, also, the data are re-represented from the form of experimental results suitable to the experiment to the form of results suitable to the theory. These are not separate reasons since the way of organizing the compressed table of results depends upon the conceptualization of the problem for form and content.

The form of the rules to be written is dictated by the form of existing rules. The computer representation for the existing rules has been described elsewhere (Buchanan and Lederberg 1971, Buchanan, Feigenbaum and Lederberg 1971). In short, a rule describing the behavior of a class of molecules in a mass spectrometer is a conditional of the form $S \rightarrow A$ where S is a description of features of molecules of the class (a 'cause') and A is a set of names of processes ('effects') which occur in the instrument. The result of applying a rule is another molecule or molecular fragment which, itself, can be processed by other conditional rules. Because this simple rule form has been easy to use and easy to modify, the program which creates new rules will create them in this form.

Transformation

The data are presented to the program as effects of unnamed actions (peaks in the FMT) for each molecule. Thus, it is necessary to infer the nature of the underlying actions if we want to put the actions into the consequent places of conditional rules. For example, the FMT of the estrogenic steroid estrone, shown in table 2, shows peaks at approximately 70 mass points (above mass 65), each of which can be identified with an elemental composition. Because of small mass differences between some groups of atoms, it is occasionally impossible to determine a unique elemental composition for a mass point within the resolving power of the instrument. In that case, the peak is identified with both (or all) possible compositions. In order to form rules explaining why those peaks appear in the data for estrone, it is necessary to transform each peak in the data to a set of possible actions which are potential explanations of the peak.

Our bias toward heuristic search has strongly influenced our choice of strategies. Generating the plausible actions and pruning with respect to the data is a prime example of heuristically searching the whole space of explanations. The space and the constraints used in generation of actions were described in section 2.

The final result of transformation is a list of peak-action set pairs for each molecule. Peaks or actions which are not a member of some pair are eliminated from further consideration in the theory formation process. Since the list of possible actions has been pruned before attempting to explain the peaks in the data, it is possible that some peaks will be unexplained. These peaks are reported by the program to provide a method of checking the validity of the pruning process. The actions which occur together in a process set as the explanation for a particular peak are redundant explanations for that peak. Table 3 shows the results for the one molecule we have been using for an example. A similar table is given for each molecule in the initial collection of molecules with which the theory formation program works.

Reorganization

The rule formation program must be able to view each action separately to determine the situation in which it is likely to occur. It must have information about the molecules for which the action does and does not occur. Therefore it is necessary to reorganize the information collected so far in this simple way.

Table 4 shows each action for which any evidence was found in the data of any of the molecules. The example molecule we have been following, estrone, is molecule number 1 in this table. As can be seen there estrone (and other molecules) shows many, but not all, of the actions in the table.

Of particular interest in the reorganization of results is the handling of redundant actions. Each action which could have been responsible for a peak in the data is given credit for that peak, and the ambiguity of the peak

(in the form of a list of the other actions which could have explained the peak) is noted. The resulting set of data is a list of actions in which each action is supported by evidence in one or more molecules. Each molecule gives the peak intensity for this action and a list of any other actions identified as being redundant explanations in this molecule. Our desire for compact presentation of results provides the rationale for a final step of reorganization, which is the compression of redundant actions into a single group of actions supported by the same evidence.

4. CHARACTERIZATION OF THE RULE FORMATION PROBLEM

The space of situations described in section 2 gives a rough idea of the nature of the rule formation problem. In Appendix 2 the rule formation problem is posed in more detail. The class of estrogens is defined thus formulating the space of molecules over which theories are to be constructed. The form of input data is presented (Appendix 2, Part A1). We present a series of problem reductions and several steps in the solution of the problem. The first step in the solution is the generation of explanatory hypotheses for the data. This is done by generating possible actions which explain the data and pruning with respect to *a priori* criteria of plausibility and *a posteriori* evidence in the data. Details of the program have been described above. An important reorganization of the data concerns leaving the explanatory mode, listing the behavior of each *molecule* individually, and approaching a generalizing mode by describing each *action* individually. In the course of doing this, the program groups together sets of ambiguous actions that are not distinguishable by the data (Appendix 2, Part A3).

In seeking to introduce generality into the explanations we ignore actions that are specific to single molecules or to a very small number of molecules in the sample. In order that we include only actions that can reasonably be applied not only to the molecules in the data, but also to any of the estrogens, we restrict our attention to skeletal actions. Section B, which develops the theory of skeletal fragmentation, begins with the definition of skeletal actions (Appendix 2, Part B1) and introduces a simple representation for estrogenic molecules using structural coding of substituents around sites. The vocabulary for specifying arbitrary subclasses of estrogenic molecules represented in the data is defined next. The primitive classes (Appendix 2, Part B3) and the non-primitive classes (Appendix 2, Part B4) are defined in a natural way. This vocabulary allows the introduction of requisite generality into the description of actions and specifying the classes of molecules that show evidence for each action. Even at this stage we have captured some explanatory power and generality into the rules that are constructed. The generality introduces predictive power to a small degree. We postpone discussion of enhancing the predictive range of rules to a later section (D) and deal with problem reduction in section C.

Because of the assumption that actions are independent of one another we do not have to treat any set of actions simultaneously (Appendix 2, Part c1). The problem then decomposes into several independent subproblems with accompanying reductions in problem size (Appendix 2, Part c2). Construction of rules then proceeds in the context of single actions. Considerations are delineated such as, which among all the non-primitive classes are suitable for associating with the occurrence of an action (Appendix 2, Part c3). The concepts of cover and honesty to data are important in this context. The space of all rules is the product space formed by taking actions from the space of actions and associating every non-primitive class of molecules with each action. The space of rules is very large. We need good pruning conditions for filtering the rules and evaluation criteria for selecting from those that remain (Appendix 2, Part c5).

The rules then need to be systematically recoded to be made applicable to a wide range of molecules in the estrogen class, rather than to narrow classes defined using substituents occurring in the sample. This enhancement of predictive power is explained in section D. The technique of term replacement is the topic in D1, and interim computer implementation is described in D2. The program Planner (Appendix 2, Part D3) is given the mandate of providing the rule formation program the information it needs to develop rules amenable to term replacement.

The rules collected for different actions now have attributes of explanation, generality and predictive power. These rules have to be codified and combined with the existing corpus of knowledge. But discussion of this aspect of theory formation is beyond the scope of this paper.

FIGURES AND TABLES

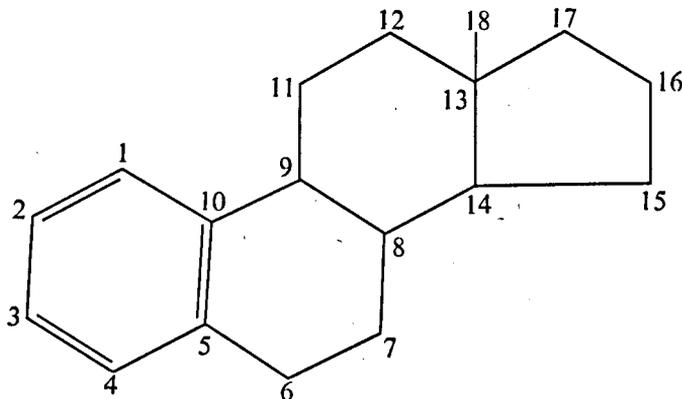


Figure 1. Estrogen skeleton $C_{18}H_{24}$ (hydrogen atoms not shown). All nodes are carbon atoms with the proper number of hydrogen attached. (Proper number = valence of atom minus number of bonds attached, e.g., 4-2 for node 15, or CH_2 .) The conventional numbering of the nodes in the estrogen skeleton is shown.

INFERENCEAL AND HEURISTIC SEARCH

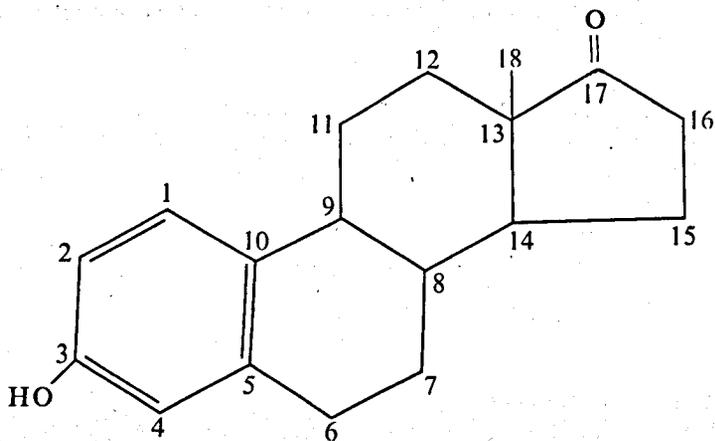


Figure 2. Structural features of estrone ((OH)₃ (O=17)). The two substituents are OH on node 3 and O= on node 17.

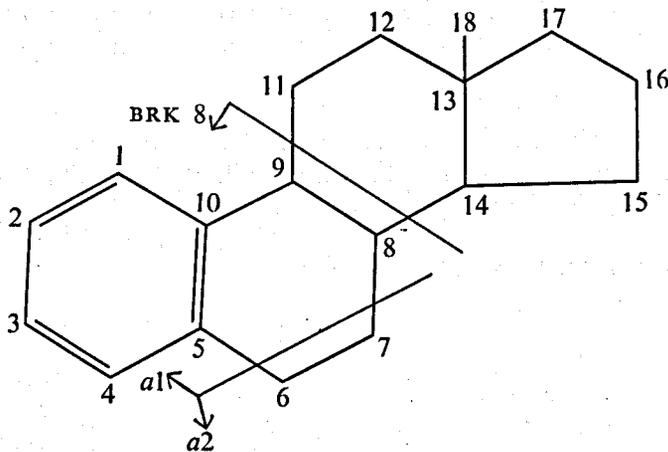


Figure 3. Example of actions in the estrogen skeleton. The arrow indicates fragment selection, the long diagonal line indicates cleavage. No migration occurs in a_1 or a_2 .

Diagram 1. Conceptualization of meta-DENDRAL.*

[DATA (Analytic Data plus structural features for each of n molecules)]



Data Interpretation: Generate actions to explain data points



[Table of Plausible Actions]



Rule Formation: Search for plausible explanations for an action A_i



Alternative explanations for each action

$$\begin{array}{l}
 S_1 \rightarrow A_1 \quad S_1 \rightarrow A_2 \quad S_1 \rightarrow A_k \\
 S_2 \rightarrow A_1 \quad S_2 \rightarrow A_2 \dots S_2 \rightarrow A_k \\
 \vdots \\
 S_n \rightarrow A_1 \quad S_n \rightarrow A_2 \quad S_n \rightarrow A_k
 \end{array}$$

* Unification of rules, the third step of the procedure, has been omitted from this discussion, and consequently from this diagram. The output of the second step, Rule Formation, is a set of rules which is a primitive sort of theory in its own right. But we recognize the convention of withholding the term 'theory' from a 'mere' collection of rules.

Table 1. Some possible cleavage sets for the estrone skeleton generated under constraints. Constraints used for generating the cleavage sets are:

- (1) Topological constraint
- (2) Chemical constraints (a-i), (a-iii), (b), (d)
- (3) Methodological constraints (a), (b).

Cleavage label	Cleavage set
BREAK 1	((13 18))
BREAK 2	((5 6) (6 7))
BREAK 3	((5 6) (7 8))
BREAK 4	((5 6) (9 10))
BREAK 5	((6 7) (7 8))
BREAK 6	((6 7) (9 10))
BREAK 7	((7 8) (9 10))
BREAK 8	((8 14) (9 11))
⋮	⋮
BREAK 33	((8 9) (9 10) (9 11))
BREAK 34	((8 9) (9 10) (11 12))
BREAK 35	((8 9) (9 10) (12 13))
BREAK 36	((8 14) (13 14) (13 17))
⋮	⋮
BREAK 65	((8 9) (9 10) (13 14) (14 15))
BREAK 66	((8 9) (9 10) (13 14) (15 16))
BREAK 67	((8 9) (9 10) (13 14) (16 17))

INFERENTIAL AND HEURISTIC SEARCH

Table 2. Analytic data (FMT) for estrone.

exact mass of fragment	Intensity	Composition of fragment					
		C	H	N	O	P	S
(6.50408085E01	1.10000	5	5	0	0	0	0)
(6.70558381E01	2.00000	5	7	0	0	0	0)
(7.70389282E01	3.00000	6	5	0	0	0	0)
(7.90540497E01	2.70000	6	7	0	0	0	0)
(8.10704498E01	1.60000	6	9	0	0	0	0)
(9.10545516E01	5.20000	7	7	0	0	0	0)
(9.20427483E01	1.20000	3	8	0	3	0	0)
(9.30702147E01	1.60000	7	9	0	0	0	0)
(9.50863121E01	1.20000	7	11	0	0	0	0)
(9.70652752E01	3.00000	6	9	0	1	0	0)
(1.03054406E02	1.20000	8	7	0	0	0	0)
(1.05069888E02	2.30000	8	9	0	0	0	0)
(1.07048971E02	3.40000	7	7	0	1	0	0)
(1.07085457E02	1.30000	8	11	0	0	0	0)
(1.08056890E02	1.10000	7	8	0	1	0	0)
(1.15054584E02	4.30000	9	7	0	0	0	0)
(1.16062028E02	1.80000	9	8	0	0	0	0)
(1.17069416E02	1.30000	9	9	0	0	0	0)
(1.19086545E02	1.10000	9	11	0	0	0	0)
.
.
.
(2.13127281E02	17.2000	15	17	0	1	0	0)
(2.14133907E02	8.30000	15	18	0	1	0	0)
(2.15137354E02	1.50000	15	19	0	1	0	0)
(2.15137354E02	1.50000	11	19	0	4	0	0)
(2.26136093E02	5.20000	16	18	0	1	0	0)
(2.27142082E02	1.40000	16	19	0	1	0	0)
(2.37128383E02	2.10000	17	17	0	1	0	0)
(2.42128541E02	3.70000	16	18	0	2	0	0)
(2.55133098E02	1.20000	17	19	0	2	0	0)
(2.68147613E02	1.30000	18	20	0	2	0	0)
(2.69155298E02	2.50000	18	21	0	2	0	0)
(2.70164935E02	100.000	18	22	0	2	0	0)
(2.71168209E02	18.5000	18	23	0	2	0	0)
(2.72172412E02	1.90000	18	24	0	2	0	0)

Table 3. Some plausible estrone actions inferred from data.

Data points		Full action labels†		
Composition of fragment	Intensity*			
C18H2402	.5	BRKO:H2		
C18H2302	.2	BRKO:H1		
C18H2202	26.1	BRKO		
C18H2102	.6	BRKO:H-1		
C18H2002	.3	BRKO:H-2		
C17H1902	.3	BRK1L		
C16H1901	.1	BRK15L:H-1		
C16H1802	1.0	BRK3L	BRK12L	BRK18L
C16H1801	1.4	BRK15L:H-2		
C15H1901	.2	BRK14L:H1		
C15H1801	1.5	BRK14L		
.	.	.		
.	.	.		
C12H1301	.3	BRK10L:H-1	BRK42L	BRK3L/47L
C12H1201	5.7	BRK10L:H-2	BRK42L:H-1	BRK3L/47L:H-1
C12H1101	1.0	BRK42L:H-2	BRK3L/47L:H-2	
C11H1301	.1	BRK1L/4H:H-2	BRK9L:H1	BRK3L/43L:H2
		BRK3L/46L:H2		
C11H1201	2.9	BRK9L	BRK3L/43L:H1	BRK3L/46L:H1

* Most of the information in this column is currently ignored since the program looks only for the presence or absence of evidence. We expect to make better use of this information, but for the present we do not.

† The action names in this column indicate the cleavages, fragment selection, and migration involved in the action. More than one action name associated with a fragment composition indicates that those actions produce indistinguishable fragments; that is, the actions are redundant explanations of the same data point. The notation for full action labels is:

full action label = BRK <action label>

action label = <cleavage label><fragment selection label>

or <cleavage label><fragment selection label><migration indicator>

or <action label>|<action label>

fragment selection label = L or H

(L indicates the fragment containing the lower-numbered node in the first bond of the cleavage set, H indicates the higher-numbered node.)

migration indicator = : H <num>

(If the migration indicator is absent, no hydrogens are transferred in the action.

If num is a negative number, that number of hydrogens migrate out of the fragment.

If num is positive, that number of hydrogen atoms migrate into the fragment.)

INFERENCEAL AND HEURISTIC SEARCH

Table 4. Partial list of processes for which evidence was found in the data for any of four estrogens.

Full action label*	Molecules which show fraction	ID	action† int.	Partial redundancies‡
ERKO:H2	2/4	4	.6	
		1	.5	
BRKO:H1	2/4	4	.5	
		1	.2	
BRKO	4/4	1	26.1	
		2	26.0	
		3	20.5	
		4	17.4	
BRKO:H-1	3/4	3	.7	
		1	.6	
		4	.3	
BRKO:H-2	4/4	2	.7	
		4	.7	
		1	.3	
		3	.1	
BRK1L:H1	2/4	3	.9	
		4	.2	
BRK1L	3/4	4	.6	
		1	.3	
		3	.3	
BRK3L	3/4	4	3.5	BRK1L/18L:H1
		1	1.0	BRK12L
		3	.6	BRK12L
BRK3L:H-1				
BRK12L:H-1				
BRK18L:H-1	1/4	3	.2	

* See explanation from table 3. The actions are clustered by groups of fully redundant actions, that is, actions which explain the same data point for all of the molecules showing the action.

† molecule 1=estrone.

‡ Partially redundant actions explain the same data point for one molecule, namely, the one whose number appears in the third column.

REFERENCES

- Buchanan, B.G. & Lederberg, J. (1971) The heuristic DENDRAL program for explaining empirical data. *Proc. IFIP Congress 71*, Ljubljana, Yugoslavia. (Also *A.I. Memo 141*, Stanford Artificial Intelligence Project. California: Stanford University.)
- Buchanan, B.G., Feigenbaum, E.A. & Lederberg, J. (1971) A heuristic programming study of theory formation in science. *Proc. Second Int. Joint Conf. on Art. Int.*, Imperial College, London. (Also *A.I. Memo 145*, Stanford Artificial Intelligence Project. California: Stanford University.)

APPENDIX 1. DESCRIPTION OF THE THEORY SPACE

1. A *Predictive Theory of Mass Spectrometry* is a total mapping from molecules to fragment mass tables where a molecule involves a description of its chemical structure; and the FMT in a simple form is a list of pairs of the form
(fragment mass . intensity of peak) – low resolution or
(fragment composition . intensity of peak) – high resolution.

It should be noted that exact mass values can be equated with fragment compositions, so that low and high resolution data differ to the accuracy (resolution) of mass measurements.

A *Predictive Theory of Estrogen Mass Spectrometry* is a total mapping from estrogenic molecules to fragment mass tables where an estrogenic molecule is a tuple of substituent radicals on an (implicit) estrogen skeleton which is shown in figure 1.

2. *Assumption*. The FMT's (mass spectra) can be explained entirely in terms of allowable *actions* of a *given molecule*.

This serves to fix instrument characteristics, instrument parameter settings and operating environment and emphasizes the repeatability of a mass spectrometric experiment.

In making this assumption one restricts oneself to a theory that may not explain the entire FMT for each given molecule, but only those portions of the data that can be explained on the basis of a well-defined set of mass spectrometric actions of the molecule. This clearly allows only approximate predictions of FMTs, leaving unexplained effects of solvents, impurities of the sample compound, and other familiar mass spectrometric effects.

Actions (defined more precisely earlier) are allowable sets of cleavages followed by fragment selection and hydrogen migration.

An *Action-based Predictive Theory* is a mapping from molecules to set of actions on molecules along with a well-defined effective procedure for predicting the FMT given a molecule and its set of actions. When the above effective procedure cannot predict the FMT but only the mass values or the composition of the fragments, without intensities, then the theory is a 0, 1 or a binary theory.

3. An *Action-based 0, 1 Predictive Theory* is a mapping and a procedure: molecule to set of allowable actions on molecule

and (molecule, actions) to table of fragment masses or of fragment compositions.

The set of actions allowable on a molecule can be specified, for example, by a predicate on the set of all defined actions on the molecule.

We will use M to denote the universe of all molecules; E to denote the universe of all estrogenic molecules; and $m \in M$ to denote a molecule in the universe; and Am to denote the set of actions definable on m .

4. A *Partial Theory* uses a predicate that is not total, that is, it is not defined over part or the whole of the domain Am . Another way to define the same is to use a three-valued total function on Am :

function on Am : for $a \in Am$, $a \rightarrow \{\text{YES, NO, DON'T KNOW}\}$

This allows the mapping on M (or E) to be partial, when the function of Am maps each action to 'DON'T KNOW'.

It is not certain whether the structural information of a molecule is sufficient to allow prediction of intensities in a fragment mass table. We wish to focus our attention for the present on 0, 1 theories. First steps toward refining the 0, 1 theories to predict intensities can be made by

(1) refining the YES, NO entries in the range of the functions to be HIGH, MODERATE, LOW and ZERO on an arbitrary scale of intensities.

(2) associating a confidence with each prediction, thus refining the DON'T KNOW entry in the range.

APPENDIX 2. DESCRIPTION OF THE SPACE OF RULES

A. Formulation of the Rule Formation task for estrogenic molecules (estrogen-meta-DENDRAL)

A1. The space of all estrogens E is defined by the space generated by allowing arbitrary substituent radicals on the kE positions around the estrogen skeleton.

$E = \{e | e \text{ is } (skeleton; sub1, sub2, \dots, sub kE)\}$

where each sub_i is any radical.

Allowable positions of substituents on estrogens:

1, 2, 3, 4, 6, 7, 8, 9, 11, 12, 14, 15, 16, 17, 18 $kE=15$

An example of one molecule of the estrogen class is estrone, shown in figure 2. It should be evident that $E \subset M$.

Let D denote the set of estrogens for which experimental data is presented, also called the *sample* ($D \subset E$). The experimental data consist of a list of molecule FMT pairs.

$DATA = \{(d, spectrum)\}$ for each d in D .

To derive illustrative examples we have used a reduced sample of 7 molecules derived from the original sample of 66. We shall refer to the latter as the reduced sample and the former as the full sample.

A2. In order to facilitate the development of an action-based theory the FMT is initially transformed to molecule-action sets, with the program described in detail in part c of this paper.

$DATA1 = \{(d, \text{actions})\}$ for each d in D .

The transformed set of data, called $DATA1$ here, is rewritten to a set ($DATA2$) in which the actions are given primary emphasis. Both these transformations are described and illustrated in part c.

$DATA2 = \{(d, (\text{action}_i, \text{fragment}_i, \text{composition}_i, \text{intensity}_i))\}$.

representing explanations of the type: In the FMT of molecule d the fragment of composition_i occurs due to action_i .

A3. Ambiguity of actions and redundancy of explanations

For two molecules mi and mj , the bonds (the action sites and consequently the actions) of mi can be put in partial 1—1 mapping into the bonds of mj , and vice versa. This is obvious for actions involving only sites on the skeleton. Such a mapping can be extended to actions involving the substituents either alone or in combination with the skeletal sites. For simplicity, we are currently generating actions for the skeleton only. Soon we expect to generate and compare actions involving substituents on the skeleton as well.

Actions related by this mapping are *corresponding* actions. An example of two corresponding actions is shown in figure 3, labeled $a1$ and $a2$.

Two actions ai and aj on a single molecule m are *partially ambiguous* for m if and only if

fragment composition for ai = fragment composition for aj .

Corollary. For partially ambiguous actions ai and aj the intensity associated with ai equals the intensity associated with aj .

Remark. For a theory that uses and predicts only low-resolution mass spectra, a weaker definition of partial ambiguity would be necessary, which involves the fragment masses instead of the fragment compositions.

Two actions ai and aj are *totally ambiguous* for D when

for every d in D such that corresponding actions ai' and aj' are defined, ai' and aj' are partially ambiguous for d .

Multiple explanations for a peak derived from ambiguous actions will be referred to as *redundant* explanations.

Totally ambiguous actions are indistinguishable within the confines of the given sample and data. They may possibly be partially disambiguated by obtaining further experimental data. Partially ambiguous actions, on the other hand, point toward a basic limitation of mass spectrometric recording. Since a fragment mass table records only fragment masses or fragment compositions and no identity of the fragment in reference to the skeleton is preserved, effects of partially ambiguous actions are not differentiable in the data. They can be differentiated only through further assumptions about actions or by resort to additional data.

B. A theory of skeletal fragmentation in estrogens, defined on substituent effects

B1. A theory of skeletal fragmentation is an action-based theory defined only over the skeletal actions As .

$$As = \text{set of actions definable on the skeleton} \\ = \bigcap_{e \in E} \{ \text{actions defined on } e \}$$

B2. Structural coding for estrogens

In the sample $D (D \subset E \subset M)$ which is a finite set of estrogenic molecules, each member is described by a finite set of substituent radical placements at specified skeletal positions. To simplify the presentation:

- (i) hydrogens attached to skeletal positions are understood when no substituent is specified for some positions.
- (ii) double bonds (or unsaturations) are simultaneously specified on pairs of adjacent positions, thus some viable pairs of positions are allowed as position specifications.
- (iii) each substituent is described only by a nominal name for present purposes, although structural information about the substituent will be made available when need arises.

Let $Fi(D)$ be the finite set of substituents (i.e. substituent labels) occurring in position i , in the molecules in D , for $i=1$ to kE

$$E' = \text{set of molecules definable using } Fi(D) \text{ for all } i \\ = \{ \text{skeleton}; f1 \in F1(D), f2 \in F2(D), \dots, fkE \in FkE(D) \}$$

The size of this space of molecules is

$$|E'| = \prod_i |Fi(D)| \quad i \text{ from } 1 \text{ to } kE.$$

By asserting that, each e' in E' represents at most one molecule we assert that we are neglecting the three-dimensional properties that discriminate molecules. E' is the intended maximum predictive domain of the theory for estrogens under coding with respect to $Fi(D)$.

Example. For a small sample of 7 molecules derived from the original 66 molecules,

- (OH 3) (O=17) (DOUBLEBOND 6 7) (DOUBLEBOND 8 9) (CH3 18)
- (OH 3) (O=17) (DOUBLEBOND 6 7) (DOUBLEBOND 8 9)
- (OH 3) (OH 11) (OH 17)
- (OH 3) (O=11) (O=17)
- (CH30 3) (O=17)
- (OH 3) (O=17)
- (OH 3) (OH 17)

$$|D| = 7 \text{ and } KE = 6;$$

- $F1(D) = \text{OH or CH30 on } 3$
- $F2(D) = \text{DOUBLEBOND or H on } 6 \text{ } 7$
- $F3(D) = \text{DOUBLEBOND or H on } 8 \text{ } 9$
- $F4(D) = \text{OH, O= or H on } 11$

$F5(D) = \text{OH or O} = \text{on } 17$

$F6(D) = \text{CH}_3 \text{ or H on } 18$ where we use H as the symbol for default values.

$|E'| = 2 \times 2 \times 2 \times 3 \times 2 \times 2 = 96$. Similarly, $|E'|$ for full sample = (approx.) 1.86 million.

B3. Primitive classes of molecules

Define S_{jk} as the set of elements of E' restricted by substituent $f_{jk} \in F_j(D)$ in position j of the skeleton, defined for j from 1 to kE and for k from 1 to $|F_j(D)|$. The primitive classes are classes defined by substituents at each position of the skeleton. There are $\sum_j |F_j(D)|$ primitive classes. Examples of

primitive classes (there are 13 for the small sample):

$S_{11} = \text{all those members of } E' \text{ having (OH 3)}$

$S_{12} = \text{all those members of } E' \text{ having (CH}_3 \text{ 3)}$

$S_{42} = \text{all those members of } E' \text{ having (O = 11)}$

B4. (Non-primitive) class of molecules

We can describe different classes of molecules by appropriately expressing them with a combination of set intersection and union operations on the primitive classes. The primitive classes thus generate a Boolean Algebra of sets, yielding a natural and convenient way of defining classes of molecules in E' .

The set in the Boolean Algebra generated by S_{jk} , define non-primitive classes in terms of membership function S_{jk} . Each molecule e' in E' can now be described in terms of the set intersection

$$S_{1x_1} \cap S_{2x_2} \cap \dots \cap S_{kE_xkE} = \bigcap_i S_{ix_i} \text{ from } 1 \text{ to } kE.$$

It should be evident that $|\bigcap_i S_{ix_i}| = 1$ and that $E' = \bigcup_i \bigcup_{x_i} S_{ix_i}$. It is also easy to verify that the Boolean Algebra is non-atomic and that the number of sets defined as such is

$$2^{|E'|}$$

thus allowing expression of any arbitrary combinations of molecules to be expressed as a class.

Example. The first molecule in the reduced sample is given by

$$S_{11} \cap S_{21} \cap S_{31} \cap S_{43} \cap S_{51} \cap S_{61}$$

There are 52 primitive classes of molecules in the full sample. It is true that, for every j

$$S_{jx_1} \cap S_{jx_2} = \emptyset$$

C. A partitioning of the Rule Formation problem

C1. Assumption. The actions associated with the molecules are independent of each other. That is, the mapping of each action a in Am for each molecule m into the set {YES, NO, DON'T KNOW} can be specified independently of the mapping for any other action a' in Am' (for the same or any other molecule).

INFERENCEAL AND HEURISTIC SEARCH

Corollary. The actions on the skeleton are independent of one another.

Remark. This independence assumption does not preclude the possibility of carrying over information learned from the efforts to solve one subproblem to another. But because rules are formulated for each action separately, it is necessary to unify rules after completion of rule formation. This independence affords a partitioning of the rule formation problem into $|As|$ subproblems each of which can be solved independently.

Let, for $a \in As$,

$$Da = \{d | d \in D \text{ and there is evidence for action } a \text{ in the spectrum of molecule } d\}$$

$$Da- = \{d | d \in D \text{ and there is no evidence for action } a \text{ in the spectrum of molecule } d\}.$$

The skeletal action formally labelled 8 is shown in figure 3.

For this action and the reduced sample,

$$D8- = \{(\text{CH3 } 18) (\text{O} = 17) (\text{OH } 3) (\text{DOUBLEBOND } 6 \ 7) \\ (\text{DOUBLEBOND } 8 \ 9), (\text{O} = 17) (\text{OH } 3) (\text{DOUBLEBOND } 6 \ 7) \\ (\text{DOUBLEBOND } 8 \ 9)\}$$

$$D8 = \{(\text{OH } 3) (\text{OH } 11) (\text{OH } 17), \\ (\text{OH } 3) (\text{O} = 11) (\text{O} = 17), \\ (\text{CH30 } 3) (\text{O} = 17), (\text{OH } 3) (\text{O} = 17), (\text{OH } 3) (\text{OH } 17)\}$$

$$|E'(D8-)| = 2 \times 1 \times 1 \times 1 \times 1 = 2$$

$$|E'(D8)| = 2 \times 3 \times 2 = 12.$$

The reduction in problem size afforded by this partitioning comes about in at least two ways: first, the rule formation for each action can be tackled separately without any information carry-over from one subproblem to another; second, the two Boolean algebras are much smaller than the Boolean algebra of the unpartitioned set. This consideration is examined below in more detail.

We can define $Fj(Da)$ and $Fj(Da-)$ on the partial data domains, and thereby define $Sjk(Da)$ and $Sjk(Da-)$ and generate the Boolean algebras $BA(Da)$ and $BA(Da-)$ to describe classes of molecules that have evidence for the action a and those that do not have such evidence. Since in general certain substituents are contributory to action a and certain other substituents inhibit action a , $Fj(Da)$ and $Fj(Da-)$ are smaller than $Fj(D)$. The size of $BA(Da)$ and $BA(Da-)$ are reduced to a mere fraction of the size of $BA(D)$. C3. We need to choose one class b from each of the Boolean algebras to define the class of molecules believed to undergo the action and the class of molecules believed not to undergo the said action. There are some prior restrictions we can impose on the choice of b from $BA(Da)$ and $BA(Da-)$. The following discussion using D is to be naturally extended to Da and $Da-$.

Define $B(D) = \{b \in BA(D) | b \cap D \text{ is non-empty}\}$

= the set of b that can be used to explain at least one molecule in the data domain.

However, since union of sets is allowable in any b , one can postulate a stronger condition on the choice of b . An element $b \in B(D)$ is said to cover or to be a *cover* when $b \cap D = D$ (or restating the same, when $b \supset D$). Define $R(D) = \{b \in B(D) | b \text{ is a cover for } D\}$. A rule for an action a then consists of two covers such that

$$b(a-) \cap Da = \emptyset \text{ and } b(a) \cap Da- = \emptyset.$$

The above two conditions can be interpreted to mean *honesty* to data, in that we require the rules to explain the data completely and without any errors. This condition may be compromised when the reliability of the data is in question or when the evidence in the data is not unambiguous.

Example.

$$B(8-) = S11 \cap S21 \cap S31 \cap S43 \cap S51 \cap (S61 \cup S62)$$

$$B(8) = (S11 \cup S12) \cap S22 \cap S32 \cap (S41 \cup S42 \cup S43) \cap (S51 \cup S52) \cap S62$$

They are respectively equivalent to the statements

$$B(8-) = (\text{OH } 3) (\text{DOUBLEBOND } 6 \ 7) (\text{DOUBLEBOND } 8 \ 9) (\text{O} = 17) \\ (\text{either CH } 3 \text{ or H on } 18)$$

and

$$B(8) = (\text{OH or CH } 30 \text{ on } 3) (\text{OH, O} = \text{ or H on } 11) (\text{O} = \text{ or OH } 17) \\ \text{and } (\text{DEFAULT VALUES at } 6 \ 7, 8 \ 9 \text{ and } 18).$$

C4. Predictions using the rules would employ the logic:

if $m \in b(a-)$ and m is not $\in b(a)$ then predict occurrence of action a ;

if $m \in b(a)$ and m is not $\in b(a-)$ then predict non-occurrence of a ;

else do not predict (i.e. predict DON'T KNOW).

Example. For (CH30 3) (OH 17) one would predict occurrence of action 8 and for (CH30 3) (DOUBLEBOND 6 7) (DOUBLEBOND 8 9) (OH 17) one would predict DON'T KNOW.

C5. The space of possible rules is very large for each action and we seek appropriate heuristics and guiding principles that can either reduce the number of alternatives to be considered or impose a simple rule for choosing alternatives.

One simple principle that is acceptable practice concerns the 'independence' of substituent effects, unless there is evidence available for 'interaction' of substituents. When a substituent is judged to have an enhancing effect on an action a , among the molecules observed, it may be readily supposed that it will have the same effect on all molecules in the domain. Our approximation to this principle involves the suggestion:

From $Sjk \cap Da$ is not \emptyset and $Sjk \cap Da- = \emptyset$ conclude $Sjk \subset b(a)$;

From $Sjk \cap Da-$ is not \emptyset and $Sjk \cap Da = \emptyset$ conclude $Sjk \subset b(a-)$;

Example. In the above example, it is reasonable to expect that any molecule having (DOUBLEBOND 6 7) will fail to show evidence for action 8 and any molecule having (OH 11) will show evidence for the same action 8.

Parsimony is often a desirable aim in constructing rules and there is a way to state preference for parsimonious rules in our formulation of the problem. A class of molecules stated as a union of disjoint subclasses is our equivalent

of a non-parsimonious class. Rules will be chosen to minimize the number of disjoint subclasses mentioned explicitly in the statement of the rule. Note that the representation of classes of molecules in terms of the primitive classes was really motivated toward this end, for each of the primitive classes is non-disjoint. When one begins to take intersections of the primitive classes and thereby refines them, then one moves away from chances of obtaining a parsimonious class.

There are hosts of other criteria that one can readily formulate that are intuitively valid to exercise, but we would like to experiment with as many as are reasonable, and explore ways of meaningfully combining criteria. The criteria in our current repertoire besides measures of simplicity/complexity of rule statements and degree of generality, include some measures based on the effort spent in formulating and validating the rules. There are criteria that involve prior knowledge of the theories of mass spectrometry and attempted carry-over of information from relevant areas of chemistry about possible effects of substituents on the behaviour of molecules. We expect to keep explicit account of the chemical knowledge that enters into the programs not only for understanding theory formation but also for understanding how changes in the knowledge base alter the result.

D. Extension of the predictive domain of the rules (planner and term-replacement)

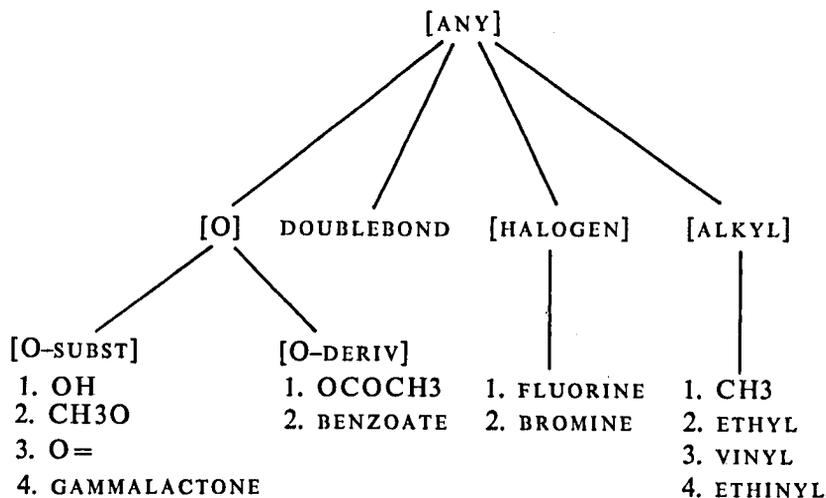
D1. The sets of rules derived for each of the actions are defined each in the narrow context of the molecules showing evidence for the action and others not showing such evidence. The rules may however be extended, as a post-process conceivably, to encompass some molecules not within the strict bounds of $E'(Da)$ and $E'(Da-)$ but within $E'(D)$. The rules also need to be extended into the infinite domain E (all estrogens) as far as possible. One possible technique of removing the bind imposed by the use of only those substituents actually seen in the data, involves the introduction of new substituent labels to serve as generalizations of substituents and effectively replacing selected terms in the rules with the generalized substituents. Such term replacement techniques raise the following questions:

- (i) when to replace terms;
- (ii) what terms to replace;
- (iii) and what terms to use for replacement?

The notion of term-replacement is based on the central assumption that the behaviour of a molecule depends solely on its topological structural features. Therefore, when two substituent radicals exhibit similar effects, one tends to relate the similarity in effects to their common structural parts. That is, extract the 'common' subgraph of the structures and extend the domain of the molecules covered to any substituent built on the common subgraph. Thus an answer to the last question raised above is:

Replace the selected part of a rule that reads ' $S_{jk} \cup S_{jl}$ ' by 'any substituent containing the subgraph common to S_{jk} and S_{jl} '.

D2. In order to spare the initial program the trouble of graph manipulation for purposes of common subgraph determination, we managed to define a hierarchy of the substituents found in our sample of 66 estrogens, by suitably constructing additional substituent labels for the generalization of substituents. The hierarchy is drawn below as a tree with the defined additional terms indicated by brackets.



Example. ($S_{41} \cup S_{42}$) which corresponds to (OH or O = on 11) can be term-replaced by (O 3) or by (O-SUBST 3) and ($S_{61} \cup S_{62}$) corresponding to (CH₃ or H on 18) can be term replaced by (ALKYL 18).

Term Definitions

O-DERIV Derivatives on any carboxylic acid.

O-SUBST Any substituent that is connected through an oxygen atom, and which is not an O-DERIV.

HALOGEN Any of Fluorine, Bromine, Iodine or Chlorine.

ALKYL Any substituent radical formed out of carbons and hydrogens where each atom is connected to another by only single bonds.

O Any substituent radical that has a leading oxygen atom (thus equivalent to either of O-DERIV and O-SUBST).

ANY Allows any substituent radical.

D3. The program to select rules out of $R(D)$ should be informed of this postprocessing, so that it may form rule-classes amenable to term-replacement. The information that this program needs in order to function effectively is to be generated by a *planning* phase, which as yet is very vaguely defined. The function of planner is to be generally described as analyzing the effects of each of the substituents with respect to each of the actions and classifying

INFERENTIAL AND HEURISTIC SEARCH

them as having an enhancing, reducing or neutral effect on the intensity of the peak corresponding to the action. It will also seek to discover those interactions between substituents that the rule formation program should consider. The suitability of two substituents to be blended by term replacement will also be assessed by planner by putting in the corresponding structural similarities of substituents and the similarity of their observed effects on intensities of peaks.

Acknowledgements

This research was supported by the Advanced Research Projects Agency (SD-183) and the National Institutes of Health (RR-612-02). Mr William C. White and Dr Dennis H. Smith provided much assistance in formulating and implementing the programs.