interpreting mass spectrometry (MS) data directly for an unknown, is restricted to class-specific rules and although it is quite useful in some domains (e.g., locating possible positions of substituents in a compound whose skeleton is known), it is not well suited to the general structure elucidation problem. The general pattern of use of mass spectral data in problems where class-specific information has not proven useful, and the compound's spectrum is not in a library, has been to use the data to determine molecular weight (or formula) with detailed structure/spectrum correlations left for retrospective rationalization. But we know that the mass spectrum contains a great deal of more specific structural information. Every ion observed is a fragment of the original molecule and because rearrangement of atoms or groups other than hydrogen is a very unfavorable process, except for certain special cases, every ion observed contains atoms which were linked together in the intact molecule. Every spectrum contains from a few to perhaps hundreds of unique ions. Even granting considerable redundancy, a spectrum should yield more useful information than is usually obtained. One approach of limited generality to extraction of structural information from a spectrum has been presented for analysis of so-called "sequence" molecules (A. Kunderd, R.B. Spencer, and W.L. Budde, **Anal. Chem.**, 43, 1086 (1971). This is a generalization of work by Biemann and McLafferty on peptide sequencing by MS. See M. Senn, R. Venkataraghavan, and F.W. McLafferty, **J. Amer. Chem. Soc.**, 88, 5593, (1966); K. Biemann, C. Cone, B. R. Webster, and G.P. Arsenault, **ibid**, 5598 (1966)). narrow category and one cannot always assign a unique structure for each of the sequence ions.

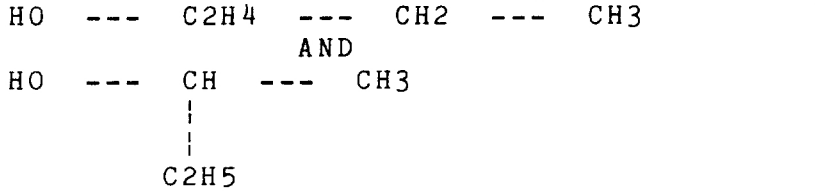
We have recently begun to explore a generation scheme which may be viewed as a generalization of sequence analysis. It makes use of a new concept called a mass distribution graph (MDG). An MDG is an entity related to (topological) chemical structures except that partitions of atoms (**e.g.**, C3H7-OCH3) are linked instead of individual atoms. Thus, one MDG may represent a whole family of topological isomers. Being a graph, it is composed of nodes interconnected by edges. Each edge in an MDG stands for one single or multiple bond, The restriction upon MDG's is that there must be some way of assembling the atoms in each partition into a connected molecular fragment (superatom) and some way of linking superatoms (using the MDG edges) into a connected chemical structure. Corresponding to each MDG is a family of structures which can be created by these two assembly steps. Within CONGEN we have the algorithms necessary for carrying out these steps.

To illustrate this, we will use a very simple example. Suppose a high resolution mass spectrum for a compound shows four major peaks, corresponding to compositions of C4H10O (M+), C3H7O (M+ - CH3), C4H9 (M+ - OH) and C2H5O (M+ - C2H5). Further suppose that the MS theory in this case is the simplest possible one; an allowed fragmentation involves the cleavage of just one single bond with no transfers of hydrogen or other neutral

species into or out of a fragment. The M+ peak defines the overall composition which, together with the MS theory, allows us to represent each remaining peak as an MDG in the form peak composition - complement, as follows:

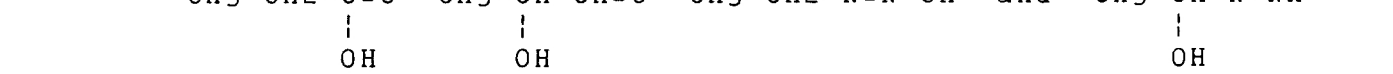
Peak	Composition	Corresponding MDG
P1	C3H7O	C3H7O ---- CH3
P2	C4H9	C4H9 ---- OH
P3	C2H5O	C2H5O ---- C2H5

The MDG generation scheme revolves around the combination of these 1-peak MDG's into more detailed MDG's each of which simultaneously accounts for several peaks. We define an "overlap operator", @, which represents this combination. Thus P1 @ P2 is the set of MDG's which have two bonds, one of which splits the overall composition according to P1 and the other according to P2. Each of these can then be "overlapped" with P3, and the resulting MDG's can be expanded into full chemical structures using structure generation and imbedding techniques already developed for CONGEN. The resulting MDG's are M1 and M2.



Only two possible structures result in this simple case, 1-butanol and 2-butanol.

MDG's can be formulated in terms of low-resolution MS data as well as high resolution data. In this case the nodes correspond to masses rather than compositions and the final expansion of MDG's to structures is accompanied by an extra step, the determination of all compositions which account for the mass of each MDG node. If the above example is treated as a low resolution problem (peaks 79[M+], 59, 57 and 45), then assuming only C, H, N and O as possible constituent atoms, six structures are possible. Aside from the above two butanols, we have:



As an initial exploration of the use of MDG's we have implemented a program, MDGGEN, which can deal with single-step fragmentations in which one or two single bonds are allowed to break, and a user-specified number of neutral hydrogen transfers are allowed into or out of the charged fragment. Because the MS theory used in MDGGEN is so simple the program has limited practical utility, but the work has demonstrated the feasibility of the MDG approach and has helped us to define the major mathematical and algorithmic advances upon which we must focus to arrive at a more general program. Two major topics are indicated.

First is the problem of formalizing the @ operator used to combine MDG's. We now have only a special-case implementation in which possible ways of overlapping MDG's based on a new data point (to yield new MDG's) were determined by hand and supplied to the program. This casewise analysis was hand tailored for the simple MS theory and a similar, though much larger, case library will be created to cover up to 3-bond, 2-step processes. These account for a great many observed peaks in typical mass spectra. The casewise approach is not sufficiently general or flexible for long term MDGGEN development, but will give us the means of creating a useful production program for short term experimentation while we explore the more general MDG overlap problem. The general solution, we believe, can be viewed as a "fuzzy" form of graph matching in which one MDG is mapped onto another with each node of the first matching either nodes or "pieces" of nodes or connected subgraphs of nodes in the second. When we have explored this concept in sufficient depth to construct an efficient, general overlapping algorithm, we will substitute it for the casewise process now in MDGGEN.

Second, we will need to increase our capacity to include constraints in the MDG generation process, constraints both on the structural features of the generated molecules and on the bonds broken in each fragmentation process. Constraints of the former type allow for the specification of desired or undesired structural features which the chemist has deduced either from other sources of structural information or from the chemical history of the unknown.

Some of this information could be incorporated at the beginning of the MDGGEN problem by defining a "starting MDG" which contains desired features. Suppose that in the above C4H10O example we had known (say, from proton NMR) that the molecule contained two methyl groups. Then rather than starting with the one-noded MDG

C4H10O

we could have started with

CH3 --- C2H4O --- CH3

and incorporated P1, P2 and P3 into this using the @ operator. The testing of substructural requirements which cannot be entered in this way (e.g., BADLIST items or overlapping GOODLIST entries) will be folded into the generation scheme wherever possible.

Constraints on the cleavage processes allow for greater precision in the specification of the MS rules which make up the theory. They will be incorporated into MDGGEN in two ways. Some constraints, primarily those concerning the allowed number of broken bonds, multiplicities of those bonds, neutral transfers and number of steps, are reflected directly in the choice of MDG structures accounting for a given peak. For example, if the simplest MS theory is used (1-bond, 1-step processes only, no

hydrogen transfers) a peak of mass p will have the MDG representation

p --- q

where q is the remaining composition. However, if single hydrogen transfers are allowed, the MDG for the peak will be any one of

p --- q p+H --- q-H p-H --- q+H

Constraints reflecting required or prohibited substructural environments for cleaved bonds will be tested for each break of each MDG as the MDG's are generated. The testing algorithms will be the ones used for structural constraints, modified to account for the fact that there are two kinds of bonds in a cleavage constraint: the "break" set, which must be tied to a set of MDG edges that corresponds to a single process, and the "ordinary" set (i.e., ordinary chemical bonds) which are free to associate either with existing MDG edges or with the implied edges inside the MDG nodes.

3.2.4 C13 NMR PLANNER

C13 NMR (CMR) is one of the most rapidly developing and potentially useful spectroscopic techniques for structure elucidation today. The chemical shift of a given carbon atom, even one which is far removed from functional groups, is sensitive to features of the local environment such as branching and steric crowding, features which are difficult to detect using other spectroscopic techniques. Yet CMR data are typically used in structure elucidation studies only to determine gross features of the carbons in the structure such as their hybridization and degree of substitution and whether they neighbor electronegative atoms, primarily nitrogen and oxygen. It should be possible to extract a great deal more structural information from a CMR spectrum automatically, and we propose to explore this possibility. Specifically, we intend to create a CMR planning program analogous to the existing PLANNER which interprets mass spectral data for unknowns.

Like PLANNER, the CMR planning program will assume a basic skeleton (a class) for the unknown and will base its analysis upon a set of class-specific CMR rules relating local environments to observed chemical shifts. There are two sources for such rules On one hand, rules have been manually extracted from CMR spectra for a variety of simple compound classes, such as acyclic alkanes, amines and alcohols, and poly-methyl cyclohexanes. These rules are available from the literature. On the other hand, our own C13 Meta-DENDRAL research proposed in a later section will be directed toward deducing from sets of spectra of known compounds relationships between local carbon environments and observed shifts. Whatever the source of the rules, the purpose of the CMR planner will be to infer from the

spectrum of an unknown the possible skeletal positions and local environments of each carbon, then to assemble full structures consistent with these possibilities. The assembly stage will also be guided by a user-supplied set of constraints similar to those in CONGEN which will allow him to enter structural information he has deduced from other sources.

In our early work on acyclic amines (39), we have demonstrated the feasibility of an automatic CMR planner for acyclic, monofunctional compounds. The research proposed here will be directed toward the much more complex problem of cyclic, polyfunctional compounds, with our primary interest being the automatic identification of polyfunctional steroids. The AMINE program, though too simple to be generalized directly to such complex cases, has given us valuable experience in dealing with CMR data, particularly in the prediction of the spectra of partially resolved structures and in the testing of such predictions against observed spectra.

Studies from both our laboratories and elsewhere have shown that in relatively rigid molecules such as steroids, CMR chemical shifts are quite sensitive to stereochemistry. This sensitivity will be reflected in the rules, and thus our structure-assembly scheme will need to include some representation of the three-dimensional features of the molecule. This will give the CMR planner the unique ability among DENDRAL programs to distinguish stereochemical isomers of a given topological structure, a step which is usually crucial to the complete solution of a structure elucidation problem. By exploring various representations of stereochemistry in the CMR planner, we expect to develop concepts which will also be useful in CONGEN and other DENDRAL programs.

3.3 New Programs for Theory Formation

3.3.1 Theory Refinement

3.3.1.1 Feedback Loops

The Meta-DENDRAL program (56) has been developed as a single pass program -- molecules and mass spectra are accepted as input data, and rules are generated in one pass through the program. A major step toward increasing the proficiency of the program is to include feedback in the control structure. A program which can notice ambiguities and uncertainties in its rule base might request a certain type of additional input (or select input from some data bank) in order to resolve discrepancies. We intend to provide the existing Meta-DENDRAL program with such abilities.

Initially, we will introduce a feedback mechanism to allow

RULEGEN to apply rules to the input data a second time (with different parameters) so that it can ignore already 'understood' data peaks and focus its attention on the more interesting 'new' data. This modification will allow experimentation with the following new strategy of rule formation:

1. Place a cutoff threshold on the intensity of input data peaks to be considered.

2. Apply existing rules (if any) to the molecule-spectrum pairs to remove 'understood' peaks from consideration. (There will be no existing rules on the first pass.)

3. Generate rules to explain peaks which are above the cutoff. Merge these into the rule base.

4. Lower the intensity cutoff threshold.

5. Go to step 2.

It is anticipated that the above strategy will focus the program's attention on the strongest unexplained peaks at each stage of rule formation. This strategy seems to parallel closely the approach taken by mass spectroscopists when analyzing data.

A second major effort to introduce feedback into the Meta-DENDRAL program will involve allowing the program to select new test data in order to

(1) increase confidence in existing rules,

(2) resolve discrepancies or ambiguities in the existing rule base, and

(3) add rules to broaden the applicability of the rule base.

In order to select new test data intelligently, the program must understand the shortcomings of the current rules. We propose to develop a formalism for concisely representing information about evidence used to support each rule. Information about possible alternate versions of the rules will allow the selection of new data to choose among competing versions of the rule. The formalism will also allow updating each rule incrementally on the basis of the correctness of each new prediction.

3.3.1.2 Alternative Representations for Rules

The rules now formed by the Meta-DENDRAL program are satisfactory codifications of the mass spectrometry processes at a given level of description. Within the model of mass

spectrometry given to it, the program finds very plausible rules. However, the success of the program is tied closely to the adequacy of the underlying model. We propose to investigate means of automatic theory formation in the absence of firm, well accepted models of the domain. The existing Meta-DENDRAL program will provide the framework for this investigation.

One way of reducing the program's dependence on a strictly defined model of the domain is to provide it with the union of terms and concepts which might plausibly contribute to explanatory rules. From this superset of terms, then, the program will be expected to select terms for rules in such a manner that the explanatory power of the rules will be maximized. Terms that contribute nothing to rules will be dropped from the model. For example, the program could discover that a potentially useful descriptive term like electronegativity is never used to explain mass spectrometry data for a class of compounds.

We can improve on the selection process by introducing a hierarchy of terms. For example, there are node and edge properties of subgraphs in a connectivity model and there are geometric properties of subgraphs in a three-dimensional model. We expect to extend the current template schema to describe hierarchies of terms and to select and reject terms from these sets.

An approach that is closer to human theory-formation methods is to give the program models of other disciplines and ask it to construct analogous models of mass spectrometry. Since some of the items in the analogy may be unnecessary when applied to mass spectrometry, the program will need to select the subset of terms that are most helpful. There is no guarantee that this method will work. But its charm lies in providing a mechanism for postulating new concepts for a domain without having to provide a generator of new concepts together with heuristics for determining their worth **a priori**. For this work on analogical reasoning, which we see as long term research, we would expect to draw largely from the model of theory formation in mathematics proposed in a forthcoming PhD thesis by Mr. Doug Lenat [Stanford University Computer Science Dept.].

3.3.2 C13 NMR Rule Formation

To extend the ideas of theory formation and test the generality of the basic concepts (56) we propose to explore a new problem domain outside of mass spectrometry. The domain of C13 NMR provides an excellent testing ground for generalization of the theory formation program since the format of the rules is significantly different from that of mass spectrometry.

C13 NMR has been characterized as the spectroscopic technique of the 1970's [69]. Our laboratories have been

involved in experimental work on C13 NMR spectra of amines, keto and hydroxy steroids (63-65). In addition, we have carried out a preliminary investigation of a Heuristic DENDRAL approach to interpretation of C13 spectra of amines [39]. Other workers have reported a related approach to the interpretation of hydrocarbon spectra [A.L. Burlingame, R.V.McPherron and D.M. Wilson Proc. Nat. Acad. Sci. USA, 70, 3419 (1973)]. Our aim in exploring C13 NMR rule formation is threefold:

1) It will greatly assist chemists who are concerned with formation of explanatory rules for C13 NMR.

2) It will be useful for assigning C13 NMR peaks in new spectra to specific carbon atoms in known structures.

3) The rules generated by Meta-DENDRAL can be used to infer structures (or partial structures) from C13 NMR data (see C13 Planner section).

There are several parallels between rule formation in mass spectrometry and C13 NMR spectrometry. In both techniques the precise reasons for molecular fragmentation (in the former) or NMR absorption (in the latter) are poorly understood. In the absence of a detailed theory capable of accurate prediction of spectra, we seek empirical rules which can relate observed data to measurable structural parameters. Some of the structural parameters presumed relevant, **e.g.**, atom type, bond multiplicities, are shared in both techniques. Some of the current Meta-DENDRAL structural manipulation functions can be used for either technique. An important difference is that the planning phase of Meta-DENDRAL (**i.e.**, INTSUM) necessary in applications in mass spectrometry is not required for C13 NMR because we will deal initially with spectra whose absorption peaks (or "shifts" relative to an internal standard) are assigned to specific atoms in the known structures. Typically scientists have sought an explanation for the C13 NMR shift of an atom in terms of the structural environment of the atom. Searching such structural environments is a problem which is amenable to solution by existing and proposed parts of the Meta-DENDRAL program.

As in applications to mass spectrometry (56) we will propose a set of factors which might affect C13 NMR absorptions. With a description of these factors we will use the Meta-DENDRAL program to produce a set of rules which will reproduce and predict resonance shifts of individual C13 atoms.

The current Meta-DENDRAL program represents a basic framework for studying C13 NMR rule formation. We believe that the program will require little revision to accommodate the differences in data and rules. We have already considered some of the problems of changing the form of rules. The subgraphs in the descriptive ("situation") parts of rules need to be expanded "outward" from a specific C13 atom instead of outward from a bond

broken in the mass spectrometer. The action parts of rules need to take account of an explicit absorption range whereas for mass spectrometry the rules predict much more precise data points (mass positions). We have made a preliminary test of the program's extensibility in the context of alkanes.

We intend to take the following steps in order to apply Meta-DENDRAL to C13 NMR data for complex molecules:

Incorporate three-dimensional relations among atoms as properties in subgraphs, in addition to connectivity and atom properties now used for rule formation. The preliminary studies on alkanes used only properties of connectivity, but we realize the necessity of describing stereochemical features of complex molecules.

Obtain a program to give us reasonable geometric models for known structures. We are currently looking at model building programs written by Wipke and Allinger [N.L. Allinger, M.T. Tribble, M.A. Miller and D.H. Wertz **J. Amer. Chem. Soc.** , 93, 1637-1648 (1971)] to see if they will fit our needs for this problem.

Study the relationship of conformation and C13 NMR shifts. We intend to start by looking at the C13 NMR spectra of simple fused ring systems in cyclohexanes and decalins, and progress toward our long-range goal of understanding the C13 NMR spectra of steroids.

3.3.3 Further Generalization of Meta-DENDRAL

One of the main motivations of this project is to develop programs and ideas that are applicable to more than a single domain. We propose to extend the generality of the Meta-DENDRAL programs to test the applicability of the knowledge-driven rule formation strategy to other data. We believe the Meta-DENDRAL strategy can be shown to be a useful complement to statistical approaches such as clustering and multiple regression.

Part of the effort of extending Meta-DENDRAL into C13 NMR rule formation will be spent on making the program general enough to work with both mass spectra and C13 NMR spectra, especially since C13 NMR spectral data accumulation is becoming rapidly a routine procedure in many organic laboratories. After this we will have a much better idea of how general our original ideas have been and what restrictions there are on the domains of applicability. A general, model-driven rule formation program will be applied to other medical and biomedical domains that will be selected for their medical relevance and their suitability for the program's development.

3.4 Applications

The attached annual report (Appendix II) summarizes our activities to date involving applications of our instrumentation resource and our programs for computer-assisted structure elucidation to chemical structure problems. These activities have included pursuit of our own mass spectrometric and structural problems, those of other members of the Department of Chemistry, collaboration with several groups in the Stanford Medical School and assistance on problems submitted by a wide variety of persons remote from Stanford who have made use of our facilities. We have so far been able to accommodate almost every request which has been made for use of the mass spectrometer and the computer programs, under the guidelines established in our current grant period.

We indicate in this section the directions we see our own interests in chemical applications taking us. On-going work with local and remote collaborators which will presumably continue into the future is also mentioned. We cannot, however, predict the kinds of applications which current or new users of our facility will bring to us. Much of the work summarized in the annual report was undertaken after informal conversations or correspondence with interested persons. We expect this to continue, we encourage it and we are taking steps (see subsequent section on increased availability) to improve our mechanisms for sharing of our resources in new applications areas.

Important research areas which we know will receive our continuing interest are the following:

3.4.1 Marine Natural Products

Professor Djerassi's laboratory is engaged in intensive structural studies of the organic constituents of marine organisms. The attached annual report (Appendix II) describes the use of DENDRAL facilities including the mass spectrometry resource and CONGEN in structural studies in this area (see also Cheer, et al. ref 59). We propose to continue these studies with special emphasis on elucidating individual structures and the sterol content of several marine organisms. We have chosen mixtures of marine sterols as candidates for computer-assisted analysis for a number of important reasons. First, not only are sterols intrinsically interesting compounds in that they are hormone precursors and important membrane constituents, but sterols derived from marine sources are particularly interesting because a number of sterols found only in marine sources possess very unusual, difficult-to-synthesize structures. These sterols are interesting not only from the standpoint of their potential biological activity and biosynthesis, but also as potential sources for starting materials in difficult steroid hormone syntheses. Second, marine sterol compositions have yielded important information which has helped clarify the phylogenetic

and evolutionary relationships among a number of classes of marine invertebrates. Evidence is now accumulating which indicates that many minor sterol constituents from marine animals are exceptionally stable molecules which have been carried intact through the complex marine food chains. A careful systematic study of marine sterols could therefore not only yield new and important compounds, but at the same time help clarify uncertain evolutionary relationships, and help disentangle complex marine food chains which are of considerable economic as well as scientific relevance. Finally, marine sterols are a fairly homogeneous class of compounds in that (1) marine sterols all possess a common nucleus which results in a number of common mass spectrometric properties; (2) marine sterols all possess very similar chromatographic properties and can be quickly and completely isolated as a single complex fraction which is amenable to a rather thorough separation and analysis by GC/MS; (3) because the fractions are generally complex mixtures (it is not uncommon for a single extract to contain upwards of 30 sterols), a great amount of time is required by highly skilled scientists in the analysis of the GC/MS data for these mixtures.

Routine analysis of the sterol content of a new mixture can be carried out with a computerized GC/MS system which includes facilities for data acquisition and reduction and subsequent library search facilities which make use of spectrum matching and GC relative retention indexes. This will quickly screen out known compounds leaving new components whose structures can be investigated further.

We propose to study new structures in a two-pronged attack using our programs for mass spectral analysis and CONGEN. Specifically, we plan to use the Meta-DENDRAL programs INTSUM, RULEGEN and RULEMOD to assist in the discovery of rules of mass spectral fragmentations to supplement available studies on the influence of side chain and skeletal unsaturation and substitution. These rules will then be used in PLANNER to assist in solving new structures. CONGEN will be used to supplement PLANNER as new features for mass spectral analysis are added to CONGEN.

3.4.2 Analysis of Organic Constituents of Body Fluids

We propose to apply our existing and proposed programs for computer-assisted structure elucidation and our GC/HRMS resource to structural problems of our collaborators in the Department of Genetics. A portion of their research is a metabolic screening program aimed at characterization of organic constituents of body fluids of patients with suspected metabolic disorders of genetic origin. (That work is funded separately under a Genetics Research Center grant, Prof. J. Lederberg, Principal Investigator.) Candidate patients are identified by collaborators of the Center grant, drawing from clinics at Stanford and other area hospitals. Urine samples, occasionally blood, cerebrospinal

fluid or amniotic fluid, are collected from these patients and turned over to Prof. Lederberg's laboratory for analysis. Analytical procedures involve chemical fractionation of the fluid into several fractions, including amino acids, organic acids and sugars. Each fraction is derivatized with appropriate reagents and subjected to GC/LRMS analysis.

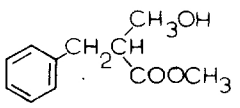
This research is in many ways ideally suited for applications of our techniques. In fact, collaboration with Prof. Lederberg's group has taken place during our current grant period, primarily involving computer programs for processing low resolution mass spectral data subsequent to data collection. Our programs for molecular ion determination and for removal of background and overlapping component interferences (CLEANUP) are part of the standard data processing procedures in Prof. Lederberg's laboratory. We also contributed some effort toward the library search facilities which are common to the mass spectrometry laboratories in Genetics and Chemistry. The analytical procedures and structural identifications in Genetics rely almost exclusively on gas chromatography and on mass spectrometric data. The CLEANUP program produces spectra which compare favorably with spectra present in our library if the compound's spectrum has been previously recorded. However, frequently new components are detected which are not present in the library. There are usually several such components in a given GC/LRMS run. Unidentified components in those experiments present important problems in structure elucidation. They can indicate metabolic abnormalities important to the future treatment of the patients. We feel our current and proposed programs and instrumentation are capable of high enough performance to provide valuable assistance in solution of these problems.

We see collaboration to make use of our facilities proceeding along the following lines: a) GC/HRMS data - empirical formulas are needed to help establish the empirical formula of the compound and of its fragment ions prior to detailed structural analysis. We can provide GC/HRMS data semi-routinely now and will be able to routinely at the outset of our proposed grant; b) CONGEN analysis - CONGEN is now capable of dealing with construction of structural possibilities. We can express many of the constraints which represent knowledge of the biochemical sources of the compounds and the chemistry of the isolation procedures. Improvements and extensions to CONGEN which we propose to implement will simplify analysis of these problems and make it much easier for the person working on a particular problem to use the program; and c) mass spectrum analysis programs - our proposed development of powerful programs for analysis of mass spectra in terms of structure includes a constructive procedure based on mass distribution graphs (see Methods Section 3.2) and the capability for testing candidate structures to determine agreement of predicted spectra with observed. Together, these developments represent a powerful amalgam with CONGEN for study of unknown structures where mass spectrometry is the primary data source.

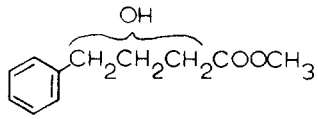
Recently we have been able to exercise some of the above procedures in the study of unknown compounds as we seek to determine where our instrumentation and programs need further attention. We outline two simple examples which are representative of the approach outlined above. We make no claim for these cases that the results could not be derived manually, but these preliminary studies indicate a strong potential for future applications.

Example 1. The patient was a mentally retarded eleven year old. The organic acid and amino acid fractions of the patient's urine revealed abundant quantities of phenylketo and phenylhydroxy acids and phenylalanine and phenylglutamine. These compounds are characteristic of phenylketonuria (PKU). Further investigation revealed the patient had been born just prior to general screening for PKU and had never been tested subsequently. The organic acid fraction contained several prominent GC peaks which were not identified by library search procedures. Subsequent chemical investigations revealed that some of the unknown GC peaks were artifactual products of the reaction of diazomethane (the derivatizing reagent) and an abundant component, phenylpyruvic acid. A GC/HRMS analysis of this fraction provided the necessary elemental composition information to begin structural analysis of the unknowns.

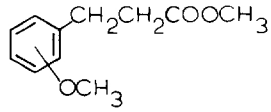
One new, non-artifactual compound, C11H1403, has been analyzed with CONGEN using a variety of constraints and structural fragments inferred from the chemical procedures, the mass spectrum and biochemical knowledge. There are nine plausible structures including branched chain phenylhydroxybutyric acids (e.g., 1) (less likely), straight chain phenylhydroxybutyric acids (2) (questionable) and o, m or p methoxyphenylpropionic acid (3) (all as methyl esters; phenolic hydroxyl groups are etherified under the derivatization conditions).



1



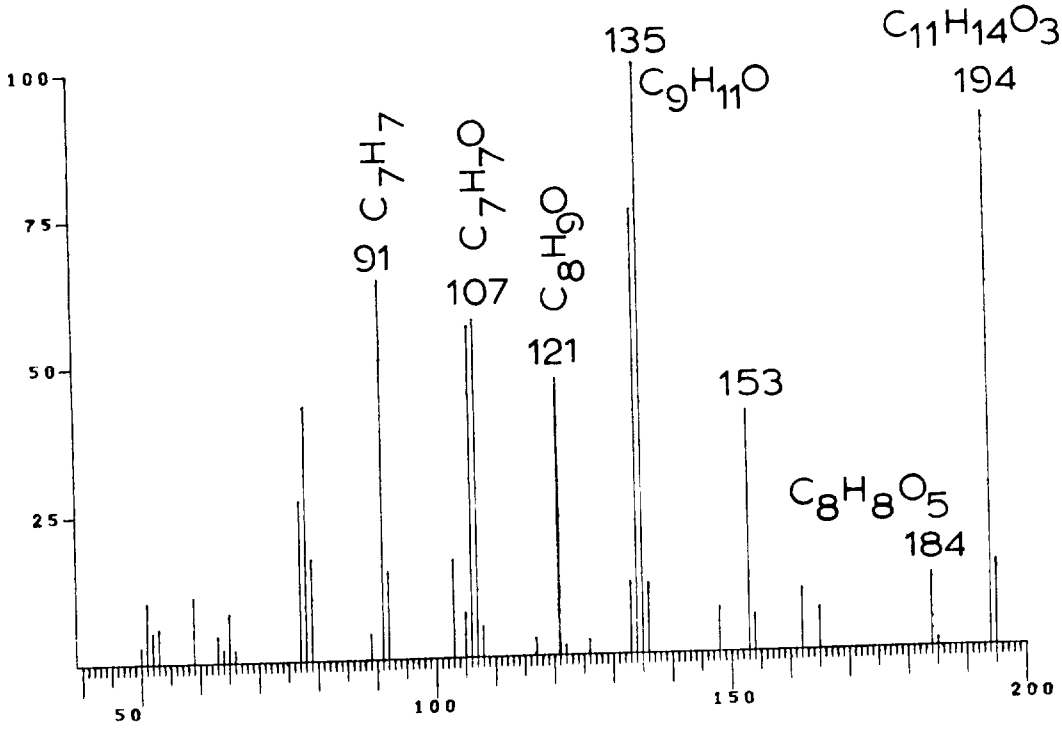
2



3

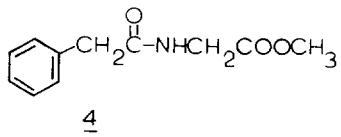
Using recently developed CONGEN functions to predict mass spectra of structures, the set of nine candidates were tested against observed elemental compositions of abundant fragment ions of mass 91 (C7H7+), 121 (C8H901+), 135 (C9H1101+) and 107 (C7H701+) (Fig. 2). Only the methoxy-substituted phenylpropionic acids (represented by 3) can yield these ions under reasonable constraints. Comparison of the spectra of authentic standards will soon be carried out to verify our hypothesis. The

Figure 2

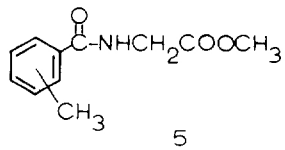

100
75
50
25
91
C_7H_7
107
C_7H_7O
121
C_8H_9O
135
$C_9H_{11}O$
153
$C_8H_8O_5$
184
$C_{11}H_{14}O_3$
194
50
100
150
200

biochemical significance of this compound remains to be assesed. Work is continuing on the structures of the artifacts resulting from the derivatization procedure.

Example 2. The urine of a mentally retarded 21 year old was subjected to the same analytical procedures. Abnormal quantities of salicyluric acid (o-hydroxybenzoylglycine), as the o-methoxymethyl ester derivative, were noted in the organic acid fraction. This compound is a metabolite of aspirin so its presence is probably not significant. However, two additional components were present in abundant quantities in this fraction. No record of them was found in our spectral library. The observed low resolution mass spectra, which share similar ions, are presented in Figure 3. GC/HRMS data revealed that the compounds are isomeric, of empirical formula $C_{11}H_{13}NO_3$. Analysis of structural possibilities with CONGEN yielded 40 structures including a variety of ways of assembling an aromatic ring, a methyl ester and an amide functionality together with two other carbon atoms. Use of mass spectrum prediction functions with a restricted theory of mass spectrometric fragmentation yielded four "most plausible" candidate structures, 4 and three isomers represented by 5.



4



5

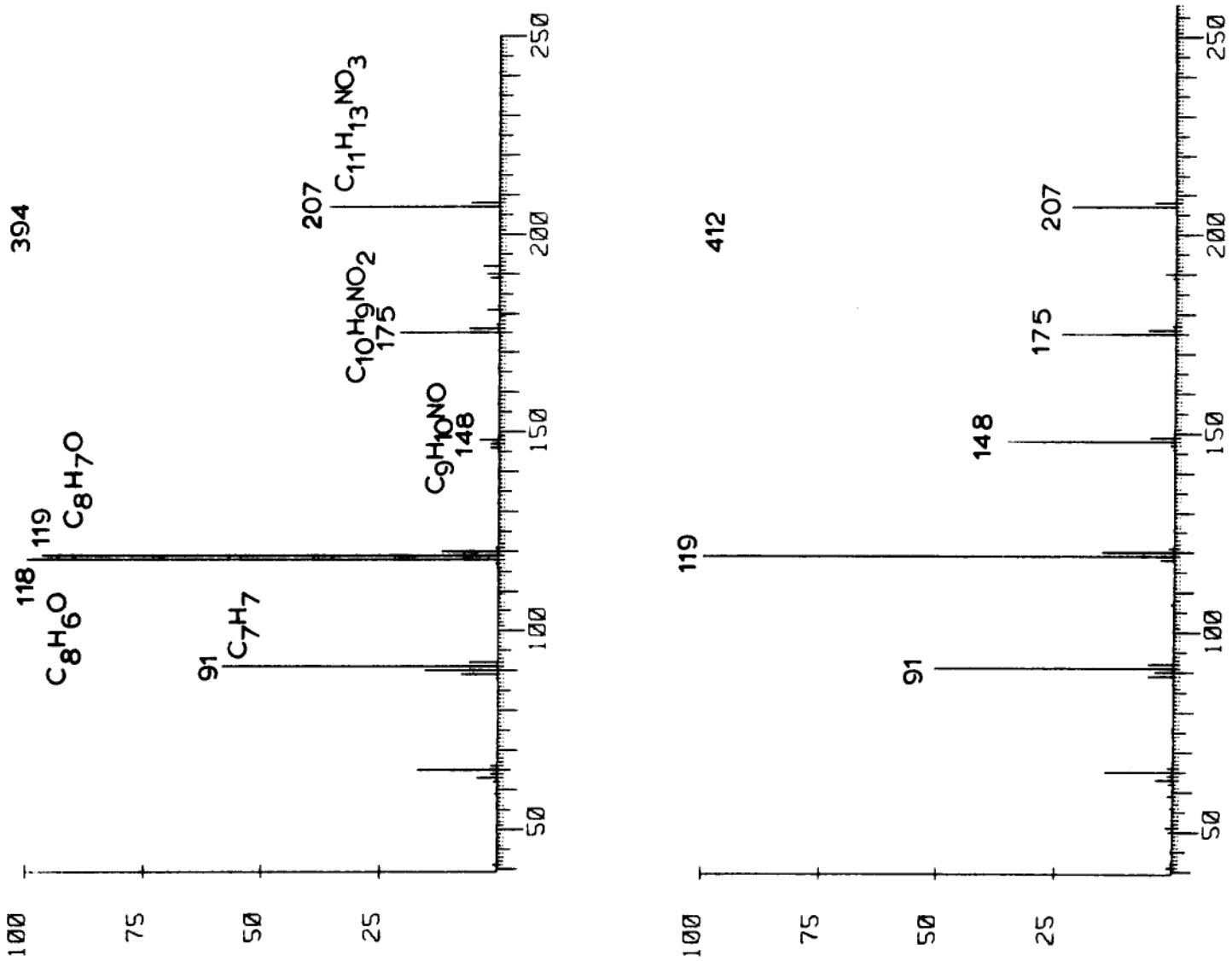
Structure 4 represents a conjugate of phenylacetic acid with glycine, and has been observed in the dog, but never in man. Structures represented by 5 are attractive because closely related isomers, which might yield similar spectra, are possible. However, there are no logical biochemical precursors for such structures. Again, we are attempting to verify our hypothesis by synthesis and comparison of spectra.

In both examples, structures which had not yet been considered by manual interpretation were derived independently by the program. In addition, other, perhaps less plausible, candidates were suggested, which gave the investigator the full set of possibilities to evaluate systematically using whatever additional knowledge or data he/she possessed.

3.4.3 Applications of Reaction Sequences

We have discussed in the Annual Report (Appendix II) our initial steps in development of extensions to CONGEN toward facilities for carrying out in the computer complex sequences of

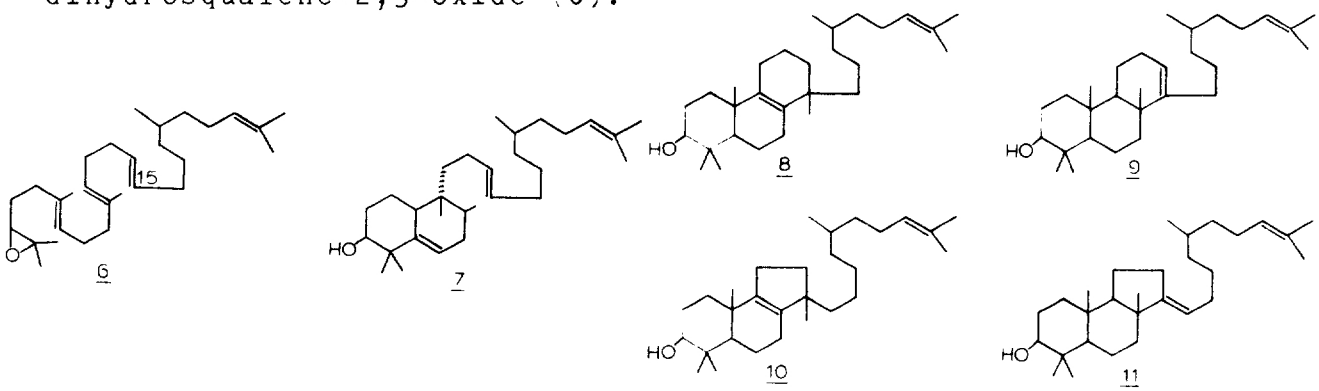
Figure 3



chemical reactions. An initial publication (ref. 61) has described the utility of this approach to structure elucidation problems and to mechanistic studies. A previous section of this proposal has described planned extensions to these facilities for studying reaction sequences.

Structural studies based on reaction sequences open up a broad class of problems of cyclizations and rearrangements to analysis with the assistance of CONGEN. Such studies do not involve assembling structural possibilities from small fragments of the molecule inferred from various data. Rather, the studies are founded on the fact that one begins with a known structure and the products must be related to the known by relatively minor perturbations of that known structure **via** a set of known reactions. We note that the ability to study reaction sequences also gives us the capability, in principle, to approach structure elucidation by hypothesizing a candidate structure and working toward a closely related solution by judicious manipulation of the candidate. We propose to explore these ideas in areas of current research interest.

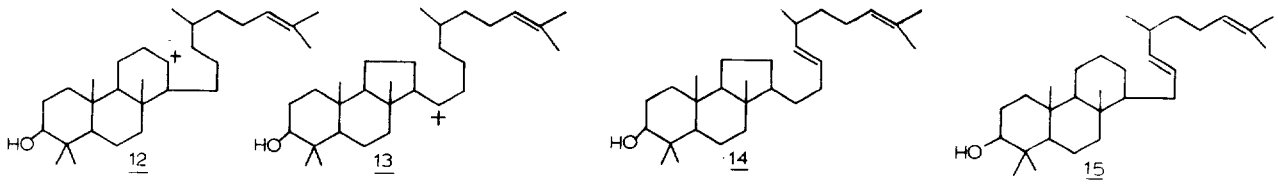
We are currently assisting Prof. van Tamelen's group in a study of unknown cyclization products using our current program. This study, described below, represents a model for an approach which we feel can be extended significantly by new developments proposed in the previous section describing the REACT program. The problem involves unknown structures from both acid and enzyme catalyzed cyclization of a squalene congener, 15'-nor-18,19-dihydrosqualene-2,3-oxide (6).



1) **Acid catalyzed cyclization of (6).** This reaction yielded a complex mixture of bi- and tricyclic alcohols. GC and liquid chromatographic analysis of the mixture yielded ten significant components. The main product was the bicyclic alcohol (7) formed in 25-30 percent yield from (6). In addition to 7, several structures possessing 6-6-5 and 6-6-6 tricyclic ring systems (ring A,B,C, of the steroid nucleus, respectively) were formed. Mass spectral and NMR data gathered on separated unknowns has led to structural suggestions for three of the components, including 8-10.

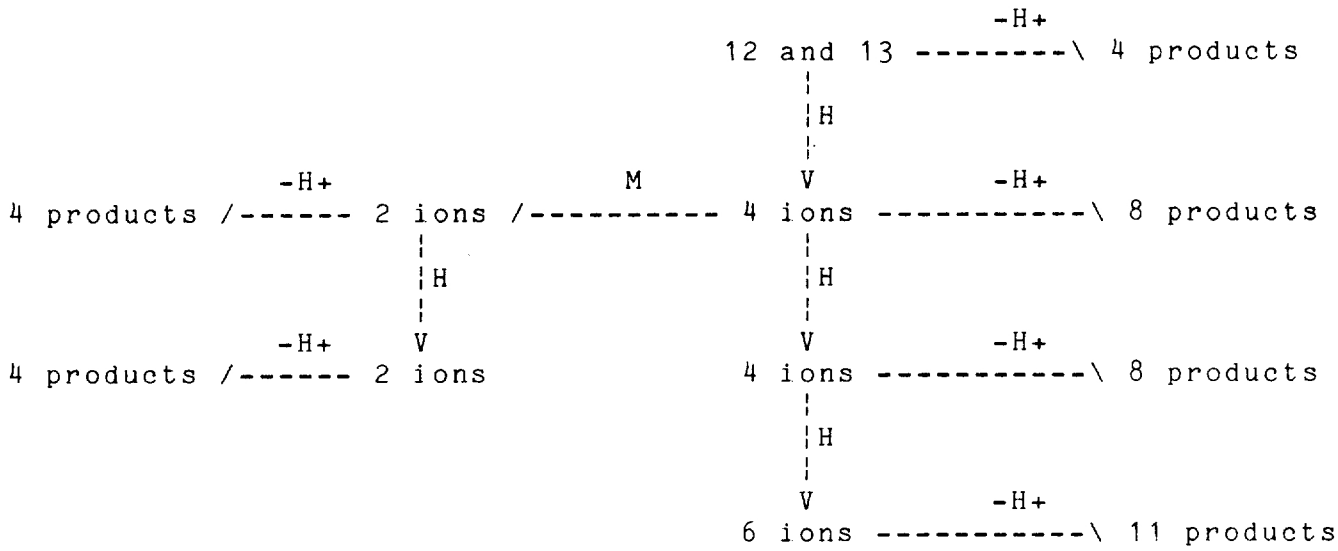
The remaining six structures remain unknowns, although structure 11 has been assigned tentatively to one compound.

We have simulated possible rearrangements of initial cyclization products to yield candidate structures for the remaining unknowns using CONGEN under a variety of constraints. Tricyclic products from such cyclizations almost always yield 6-6-5 or 6-6-6 ring systems. This constraint was used to postulate two tricyclic carbonium ions as starting points for further rearrangement (12 and 13). These carbonium ions were allowed to rearrange via 1,2 shifts of hydrogen atoms or methyl groups, with the terminating condition of loss of H+ to yield the observed double bond (all structures are thus tricyclic and possess two additional degrees of unsaturation as two double bonds).



Shifts were allowed only when the resulting carbonium ion possessed the same or higher degree of substitution as its precursor. Collection of products after each step of the following sequence yielded a total of 17 unique final structures, including 8 - 11. The other 13 candidates are under investigation as possible structures for the remaining unknowns.

Rearrangement Processes



Key: -H+ means loss of H+ to yield a double bond in the product.
H means 1,2 hydrogen shift.
M means 1,2 methyl shift.
There are only 17 unique structures; the same structure ca be produced in different steps.

2) **Enzymatic Cyclization of 6.** Incubation of 6 with a squalene oxide-lanosterol cyclase preparations obtained from microsomes of rabbit livers yielded several products. One major product was purified by chromatographic techniques and analyzed by mass and NMR spectrometry. The empirical formula was the expected C29H500 and spectral data indicated a tricyclic system.

The unknown was subjected to oxidative cleavage with OsO4/NaIO4 to help locate the positions of the double bond. Spectral data collected on the product were strongly suggestive of an aldehyde of molecular formula C20H3402, implying loss of a C-9 unit in the oxidative cleavage. This was accompanied by loss of another degree of unsaturation, implying loss of nine terminal atoms in the side chain.

Of the 17 structures from the above simulation of rearrangement processes, only one, 14, has a double bond in a position which would yield a product consistent with the observed data. This structure is an alternative to a manually derived possibility, 15, which necessarily must arise from a more complex rearrangement process. Given these two candidates (14 and 15) it is possible to design experiments to differentiate between them. Further work is now being carried out to solve this problem.

3.4.4 C13 NMR Applications

C13 NMR has been a topic of interest in Professor Djerassi's laboratory for several years and also a subject of some earlier DENDRAL work. Experimental data have been collected for amines, keto steroids and hydroxy steroids (63-65). A computer program was written here [39] which used a set of predictive rules to deduce the structures of acyclic amines. Presently polyhydroxy-steroids are of primary concern in Prof. Djerassi's labs. The formation of rules for the hydroxy-steroids should be an easier task than for the keto-steroids due to the greater structural distortions of the steroid skeleton caused by the latter. A set of rules for the hydroxy-steroids could always be used in the analysis of keto-steroids since these compounds can be chemically reduced to the hydroxy-steroid and analyzed as such. Thus a set of rules for the hydroxy steroids will assist in the analysis of two classes of steroids. A summary of the hydroxy steroid data was given by Eggert (65) which pointed out trends in the data. Presently further studies are being made to assess the effects of steric crowding and skeletal distortions upon the C13 chemical shifts. These studies will aid the Meta-DENDRAL program in the selection of the terms which should contribute to the rule description. This work is being supported in part by NIH grant AM17896.

3.5 Increased Availability

3.5.1 Continued Collaboration and Solicitation of New Efforts

The DENDRAL project, one of the major users of the SUMEX-AIM computer facility, has formed a small community of regular, remote users. This "exodendral" community has continued to provide valuable contributions to program development, although the growth of this community has had to be slowed in response to increasing demands by other projects upon the SUMEX-AIM facility. As an example, for the months of September 1975 to February 1976, the number of CPU hours used by exodendral persons amounted to at least 8 percent of the CPU hours used by the DENDRAL project. There are currently four remote researchers whose groups regularly use CONGEN in their day to day work. Additionally,

there are several remote users who use their accounts on an occasional basis, or who access SUMEX-AIM via the GUEST mechanism.

There have been several applications of our GC/HRMS resource and CONGEN to structural problems of other members of the Stanford community and researchers both in the U.S. and abroad. These collaborations are summarized in the attached annual report (Appendix II). Contacts were made with these people in a variety of ways. We have actively encouraged persons engaged in structure elucidation to consider use of our CONGEN programs (e.g., Drs. Karliner, Nakanishi, Minale). Usually this has involved solution of a previously solved problem to indicate capabilities and limitations, followed by further collaboration on new problems. Informal discussions among scientists at meetings have inspired new applications. We have, in all of our publications, announced that our facilities were available to an outside community of users within the limits of available resources.

We feel our efforts have been successful in encouraging new researchers and solving important problems. To date we have not had to deny use of our facilities to anyone who came to us with a resonable request. We are currently facing problems of high computer system loading on SUMEX. This restricts the utility of an interactive program like CONGEN due to slow response time. We have requested that people compute in off hours and have added facilities to CONGEN to make this easy to do. Within these resource limitations, we plan to continue existing collaborative efforts and to solicit new collaborators as we have done the past year of our current grant. These collaborations have been an immense benefit in improving our GC/HRMS resource and CONGEN. Our facilities have had to confront real-world problems with all the attendant uncertainties and assumptions characteristic of such problems. This experience has provided the background for our future plans in making our facilities more widely available. With a growing user community, additional burdens are placed on the GC/HRMS and SUMEX resources. Some of our proposed new program developments are directed specifically to easing these burdens. There are, however, several additional ways of increasing availability, especially of CONGEN and new extensions to it, which are discussed in the subsequent section. Recent applications of our programs to structural problems of our collaborators are summarized in Appendix II, Section 3.4.

3.5.2 Program Translation

Translation of CONGEN

Although it has proven to be a useful research tool for chemists, CONGEN is considerably larger and more time-consuming than it could be. Its development has been an evolution involving the work of many people over several years, and most of

it is written in INTERLISP, a language which promotes rapid program development but which is not noted for its run-time efficiency. In retrospect we feel that this was the proper course - we could not have reached the current level of complexity and sophistication in CONGEN otherwise - but the result is a production-level program which, because it was not designed as a single, efficient package, is rather wasteful of computer resources.

Because of the demands which CONGEN places upon SUMEX, we probably will not be able to offer use of the program to all who have fruitful applications. Not only does this deprive the chemical community of a useful tool, but it limits the number and variety of new applications to guide us in further program developments. We propose to ease this problem by recoding CONGEN in a more efficient and exportable computer language. Greater efficiency will increase CONGEN's productivity, allowing us to offer the resource to more users at SUMEX, while exportability will allow others to transfer the program to their facilities, relieving SUMEX of the burden of supplying access for routine, non-developmental use.

The language chosen for the translation is the ALGOL subset of SAIL. This choice was made for four reasons. First, an ALGOL-like language is preferable to FORTRAN because the former allows recursive programming techniques to be used. It would be possible to rework CONGEN in terms of non-recursive algorithms, but recursion is such an integral part of the logic of the program that such a transformation would be quite difficult. Second, although probably not as efficient as FORTRAN, the SAIL compiler creates reasonably fast and compact machine code. We have done some experimental translations of a few key segments of CONGEN and find a 10- to 15-fold improvement in running time, an improvement which will greatly ease the impact of CONGEN on SUMEX. Thirdly, SAIL is designed for the standard TOPS-10 operating system on the PDP-10, a fairly common research computer configuration accessible to a large number of chemists at both universities and industrial research facilities. We believe that such outside users will have relatively little difficulty mounting a SAIL version of CONGEN on their local facilities. Finally, compared to LISP, ALGOL is a more widely known language by itself or as the basis for other languages such as PL/1, PASCAL and SIMULA. In the ALGOL subset of SAIL, CONGEN will be significantly easier both to understand and to modify by interested non-Stanford workers.

One other reason for selecting SAIL warrants special mention. A proposal has recently been submitted as an extension of the SUMEX grant to develop a machine-independent language called MAINSAIL which, in many respects, is quite close to SAIL. Particularly, the subset which we will be using for CONGEN is virtually identical between the two languages, and transferring CONGEN from SAIL to MAINSAIL would not be a major task. One design criterion of MAINSAIL is transferability from one type of

computer to another - all that is needed for a new machine is a "MAINSAIL bootstrap" package to define basic machine operations and input-output characteristics. A preliminary version is now available for the PDP-10 under both the TENEX and TOPS-10 operating systems, and for the PDP 11/45 under the RT-11 system. A bootstrap package is being designed for ORVYL, the local time-sharing monitor for the IBM 370/168, as well. Although we are not specifically proposing the coding of CONGEN in MAINSAIL because the funding of the MAINSAIL effort is not yet certain, we are aware of that effort and will maximize MAINSAIL compatibility as we proceed with the CONGEN translation. When MAINSAIL matures to a stable and widely-available language, we feel that it will provide the ideal mechanism for implementing CONGEN on a variety of other machines including smaller laboratory systems such as the 11/45.

There are two existing facilities which will ease the translation and will enable us to reach a workable balance between the run-time efficiency of SAIL and the program-development aids of INTERLISP. First, the structure of the TENEX operating system allows one to run simultaneously two or more sections of a program written in different languages, with communication between them taking place through a shared file or a shared segment of memory. This means that not all of CONGEN needs to be translated at once. Rather, it can be transferred a piece at a time from INTERLISP to SAIL. Not only will this ease the problems of debugging a large and complex system, but it will allow us to retain the more rapidly-changing developmental portions of CONGEN in INTERLISP for as long as possible as the more stable sections are translated. Even when all of CONGEN has been transferred to SAIL, we expect to maintain a SAIL-INTERLISP interface so that new ideas may be tested easily in the latter. The prototype for this "pipeline" between the two languages already exists in the linkage between CONGEN and the SAIL program responsible for fragment imbedding and structure canonicalization. The SAIL segment contains a monitor program plus a set of modules which the monitor can call. The INTERLISP segment passes data and control information to the monitor, and collects output from it. We will retain this structure so that even when essentially all of the control is given to the SAIL portion, it will still be possible to "call" INTERLISP for specialized or experimental types of processing. Of course, this mechanism will not be used in any export version of CONGEN, but it will substantially enhance the flexibility of the system for local research.

The second existing facility is a cross-compiler we have developed which translates a specialized ALGOL-like subset of INTERLISP into SAIL. The subset, called SAILISP, can be used to create and test ALGOL programs in the highly interactive and well engineered environment of INTERLISP. Once a program or portion thereof has been perfected in SAILISP, the cross-compiler is used to translate it automatically into SAIL code which can be compiled and run in the normal fashion. Though SAILISP does not

provide easy processing of linked lists, a central concept in the LISP language, it does allow a programmer to build a system interactively in small pieces, debugging and modifying each piece using the powerful INTERLISP editor and error handling package. We have found that the ease of programming in INTERLISP results as much from these interactive aids as it does from the basic structure of the language itself, and SAILISP makes these aids available for SAIL programming. In conjunction with the SAIL-INTERLISP interface described above, SAILISP will provide a well balanced system not only during the recoding of the existing algorithms, but for future CONGEN research.

3.5.3 GC/HRMS System

We will continue to run samples under our current guidelines which stress that the facility is to be used for important structural problems of biomedical relevance, but not for obtaining routine mass spectra from crude reaction mixtures. Within these guidelines we have been able to entertain nearly all requests for spectra while continuing our active program of instrument and program development. As this development requires less and less instrument and computer time, additional time will be available for obtaining high resolution and GC/HR mass spectra. We are already taking advantage of this available time in our own research on marine natural products and our collaborations with local persons at Stanford. We should have more flexibility in the future, however, and we will encourage our remote collaborators to make use of our facilities for GC/HRMS to help solve their structural problems.

3.6 The GC/HRMS Resource

In previous sections we discussed the use of the GC/HRMS resource as a tool to provide necessary data for our structural studies. We also discussed the probable increased availability of the system as time required for development decreases. We propose to devote our attention to maintenance of the system and development of a detailed understanding of its performance in a variety of applications. We also propose some further developments to improve the sensitivity and throughput of the system.

Although maintenance of a system may seem trivial, in fact maintenance goes far beyond actually keeping all the parts in working order. It means having a trained operator who can take precautionary measures to avoid down time and who can recognize when performance is deteriorating, however slightly. It means devotion of significant programmer time to carry out modifications to existing software because new chemical problems frequently require new data reduction techniques.

The developments we propose are simple in concept but are potentially very valuable. They are described in the following subsections.

3.6.1 Increased Sensitivity

We propose to develop the data reduction tools required to scan spectra at lower resolving powers. We know from past studies (A.L. Burlingame, D.H. Smith, T.O. Merren, and R.W. Olsen, in "Computers in Analytical Chemistry," (Vol. 4 in Progress in Analytical Chemistry series), C. H. Om and J. Norris, Eds., Planum Press, New York, N.Y., 1970, p. 17) that mass measurement accuracy (and thus the certainty with which elemental compositions can be assigned) decreases only slightly in scanning at lower resolving powers. The sensitivity change in reducing resolving power can be dramatic, at least a factor of ten in going from a resolving power of 10,000 to 1,000 on the Varian-MAT 711. Obviously, it would be better to operate the instrument at lower resolving powers, except that problems arise because spectral peaks which were resolved at high resolving powers may overlap at low resolving powers. We routinely operate at resolving powers of 4,000 to 5,000 in GC/HRMS mode. We have found it necessary even at these moderate resolutions to implement a scheme for doublet resolution (see Appendix II, Annual Report, for a detailed description) to separate ions from the reference compound from those of GC column bleed and the sample. This approach has generally proven sufficient because in most of our applications, overlapping triplets or higher multiplets of ions are unlikely. At lower resolving powers, however, we know that the simple doublet resolver will be insufficient. Therefore, we propose to implement a multiplet resolver effectively to restore some of the resolution lost by the mass spectrometer.

Multiplet resolution techniques applied to mass spectral and many other types of data have been reported for years. We propose a seemingly minor, but critical, twist to these procedures, namely, using a peak model based on measurement of actual mass spectral peaks immediately previous in the scan as the basis for performing this resolution. Drawbacks to multiplet resolution procedures include the facts that they are time consuming and that almost every procedure employs an assumed "ideal" peak shape. We can do nothing about the extra time required for data reduction, but we think the increased sensitivity more than justifies it. But we have found in all our efforts toward evaluation of instrument performance and doublet resolution that an accurate and reliable system must be based on the measured performance (e.g., peak shape, dynamic resolution, etc.) of the mass spectrometer, not an idealized model. Thus, we will use a peak model based on measured peaks which are presumed singlets as the basis for multiplet resolution.

3.6.2 User Interface

We will improve the facilities for examining the large volumes of data produced in a GC/HRMS experiment so that the person whose sample was run can explore his own results. We have many of the file handling routines to recover easily various experimental results and the display routines to display on a CRT or produce on a hard copy plotter any of a variety of results which can be derived from a scan from calibration data to final assigned elemental compositions. We propose to provide simple procedures for examining these data, doing library searches and performing inter-experiment comparisons of results. This will increase the throughput of the laboratory because the examination of data can be done in the off hours, leaving more prime time available for running additional samples.

4 BIBLIOGRAPHY

DENDRAL PUBLICATIONS

(1) J. Lederberg, "DENDRAL-64 - A System for Computer Construction, Enumeration and Notation of Organic Molecules as Tree Structures and Cyclic Graphs", (technical reports to NASA, also available from the author and summarized in (12)). (1a) Part I. Notational algorithm for tree structures (1964) CR.57029 (1b) Part II. Topology of cyclic graphs (1965) CR.68898 (1c) Part III. Complete chemical graphs; embedding rings in trees (1969)

(2) J. Lederberg, "Computation of Molecular Formulas for Mass Spectrometry", Holden-Day, Inc. (1964).

(3) J. Lederberg, "Topological Mapping of Organic Molecules", Proc. Nat. Acad. Sci., 53:1, January 1965, pp. 134-139.

(4) J. Lederberg, "Systematics of organic molecules, graph topology and Hamilton circuits. A general outline of the DENDRAL system." NASA CR-48899 (1965)

(5) J. Lederberg, "Hamilton Circuits of Convex Trivalent Polyhedra (up to 18 vertices), Am. Math. Monthly, May 1967.

(6) G. L. Sutherland, "DENDRAL - A Computer Program for Generating and Filtering Chemical Structures", Stanford Artificial Intelligence Project Memo No. 49, February 1967.

(7) J. Lederberg and E. A. Feigenbaum, "Mechanization of Inductive Inference in Organic Chemistry", in B. Kleinmuntz (ed) Formal Representations for Human Judgment, (Wiley, 1968) (also Stanford Artificial Intelligence Project Memo No. 54, August 1967).

(8) J. Lederberg, "Online computation of molecular formulas from mass number." NASA CR-94977 (1968)

(9) E. A. Feigenbaum and B. G. Buchanan, "Heuristic DENDRAL: A Program for Generating Explanatory Hypotheses in Organic Chemistry", in Proceedings, Hawaii International Conference on System Sciences, B. K. Kinariwala and F. F. Kuo (eds), University of Hawaii Press, 1968.

(10) 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, July 1968).

(11) E. A. Feigenbaum, "Artificial Intelligence: Themes in the Second Decade". In Final Supplement to Proceedings of the IFIP68 International Congress, Edinburgh, August 1968 (also Stanford Artificial Intelligence Project Memo No. 67, August 1968).

(12) J. Lederberg, "Topology of Molecules", in The Mathematical Sciences - A Collection of Essays, (ed.) Committee on Support of Research in the Mathematical Sciences (COSRIMS), National Academy of Sciences - National Research Council, M.I.T. Press, (1969), pp. 37-51.

(13) G. Sutherland, "Heuristic DENDRAL: A Family of LISP Programs", Stanford Artificial Intelligence Project Memo No. 80, March 1969.

(14) 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.

(15) A. M. Duffield, A. V. Robertson, C. Djerassi, B. G. Buchanan, G. L. Sutherland, E. A. Feigenbaum, and J. Lederberg, "Application of Artificial Intelligence for Chemical Inference II. Interpretation of Low Resolution Mass Spectra of Ketones". Journal of the American Chemical Society, 91:11 (May 21, 1969).

(16) B. G. Buchanan, G. L. Sutherland, 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 (1970), (also Stanford Artificial Intelligence Project Memo No. 99, September 1969).

(17) J. Lederberg, G. L. Sutherland, B. G. Buchanan, and E. A. Feigenbaum, "A Heuristic Program for Solving a Scientific Inference Problem: Summary of Motivation and Implementation", Stanford Artificial Intelligence Project Memo No. 104, November 1969.

(18) C. W. Churchman and B. G. Buchanan, "On the Design of Inductive Systems: Some Philosophical Problems". British Journal for the Philosophy of Science, 20 (1969), pp. 311-323.

(19) 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.

(20) 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).

(21) Y.M. Sheikh, A. Buchs, A.B. Delfino, G. Schroll, A.M. Duffield, C. Djerassi, B.G. Buchanan, G.L. Sutherland, E.A. Feigenbaum and J. Lederberg, "Applications of Artificial Intelligence for Chemical Inference V. An Approach to the Computer Generation of Cyclic Structures. Differentiation Between All the Possible Isomeric Ketones of Composition C6H10O", Organic Mass Spectrometry, 4, 493 (1970).

(22) 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", Helvetica Chimica Acta, 53, 1394 (1970).

(23) 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 (1971). (Also Stanford Artificial Intelligence Project Memo No. 131.)

(24) A. Buchs, A.B. Delfino, C. Djerassi, A.M. Duffield, B.G.