DATE: October 2, 1973

TO : J. Lederberg
C. Djerassi
E. Feigenbaum

FROM : B. Buchanan, D. Smith,
T. Rindfleisch A. Duffield

SUBJECT: DENDRAL Goals and Mike Oxman's Visit

The attached outline of long-term and short-term goals for Parts A & C covers research items that we believe are feasible. We do not know how well they fit with NIH goals. Therefore, if appropriate, we would like to find a tactful way of getting Mike Oxman's reactions to these items.

We are prepared to offer the available software as a service to others (e.g., on SUMEX) in support of specific resource activities, if that is desirable. Making one set of programs accessible to a large community is preferable to exporting copies of those programs.

In short, we hope Mike's visit will help us arrive at a set of guidelines for writing the DENDRAL renewal proposal.

k/
Att.

9 AM. EAF.
12 lunch. - Faculty Club.

Fri 1/3/42

September 28, 1973

DENDRAL Renewal: Potential Research Topics

PART A: Near-term

I) PLANNER Improvements.
   a) Better mechanism for input of rules for classes of compounds with simple superatoms (e.g. -N-).
   b) Incorporate existing, extended molecular ion determination routine

II) PLANNER Utilization
   a) Choose areas in which program is competent. Compound class known or severely restricted; mass spectral fragmentation rules known.
   b) Suggest - Marine sterols; juvenile hormones; analysis of specific bodily fluid components in support of shared resource Part B; other classes of steroids.

III) Structure Generator
   a) Develop an interactive version of complete Generator for chemists' use.
   b) Apply complete Generator to problems for which it is suited, i.e., where necessary constraints may be supplied by interactive guidance; e.g., isomer interconversion problems, labelling problems.
   c) Explore in depth the problems of constraint implementation; determine if existing algorithm is suitable, or new approaches are required.
   d) Continue development of ancillary support for the Structure Generator: extensions to CATALOG; enumeration (counting) algorithm for verification; carbocyclic ring index; PLUME catalog; non-planar graph catalog.

PART A: Long-term

The activities summarized in the following sections are directed toward development of software systems to attack the general problem of computer-aided or computer-directed molecular structure elucidation. The heuristic search paradigm as embodied in the plan-generate-test strategy is not only the most elegant approach, but is necessary for confidence in answers (thoroughness) and, besides, we know how to do it.

I) PLAN
   a) Chemists' inferences - an experienced chemist is capable of excellent structural inferences from diverse spectroscopic data. Initial efforts would concentrate on chemists' planning rules (constraints) coupled to an interactive generator.
   b) Program inferences - develop and study the performance of programs designed, like the planner for mass spectrometry, to develop structural inferences automatically from other types of spectroscopic data.
   c) Explore planning strategies to coordinate planning based on data from different sources interpreted by different experts, one per source.

II) GENERATE
- a) Develop a structure generator which is knowledgable about chemistry; i.e., a generator which is designed with the types of constraints which must be applied in mind.
- b) Develop a sophisticated chemists' interface to the generator (presumably interactive) which allows input information in chemical "language" independent of the inner workings of the generator.
- c) Theta-DENDRAL - based on a particular problem, knowledge of constraints and knowledge of the generator, develop a strategy for solving the problem.

III) TEST
- a) Extend existing Predictor to other spectroscopic techniques.
- b) Develop the capability for examining lists of candidate solutions to determine how they differ.
- c) Coupled with (b), suggest experiments for differentiation of groups of structures.

PART C: Near-term

I) INTSUM Extensions
   a) Analyze summaries of mass spectra with respect to known mechanisms, such as alpha-cleavage, in order to increase immediate utility.
   b) Develop an interactive version of the theory formation program that will answer specific questions about proposed new rules.

II) INTSUM Utilization
   a) Interpret and summarize the mass spectra of important compounds for which no theory exists.
   b) Confirm existing theory before adding it to the performance program.

III) Bond Environment Analysis
   Analyze the bond sites of molecules to determine common features influencing fragmentation.

PART C: Long-term

I) Mass Spectrometry Theory
   a) Develop capabilities for automatically modifying the performance program's existing mass spectrometry theory.
   b) Develop the capability of selecting the level of theory within which rules best explain the data.

II) Extensions of Theory Formation Ideas
   a) C^{13} NMR - develop a set of rules for interpreting C13 NMR data in much the same way as we have for mass spectrometry data.
   b) Program Writing - extend the ideas to the domain of computer programming. For example, find features of programs that cause common problems or find a set of productions that 'explain' a given set of input/output pairs.