

STANFORD UNIVERSITY  
STANFORD, CALIFORNIA 94305

DEPARTMENT OF CHEMISTRY

November 5, 1976

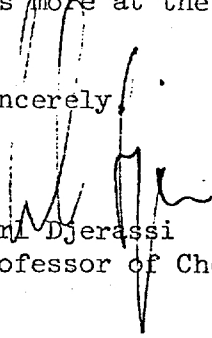
Dr. S. Stephen Schiaffino  
Office of the Assoc. Dir. for Scientific Review  
Division of Research Grants  
NIH  
Bethesda, Maryland 20014

Dear Dr. Schiaffino:

Enclosed is our response to several of the questions Dr. Lipkin indicated were not clearly answered in the DENDRAL proposal to NIH. Thank you for the opportunity to provide more information on the proposal to the site visiting committee.

We are certainly willing to discuss these points more at the site visit, but we hope this material is helpful.

Sincerely,



Carl Djerassi  
Professor of Chemistry

CD:jk

DENDRAL: Resource-related Research in Computers and Chemistry  
(RR-00612)

Principal Investigator: Carl Djerassi  
Department of Chemistry  
Stanford University

The following material is supplied in response to several questions raised by Dr. Bernice Lipkin, Executive Secretary of the Computer and Biomathematics Study Section in preparation for a site visit review of our competing renewal application submitted June 1, 1976.

1. To what resource is DENDRAL research related?

The resource to which the DENDRAL project has been related since initial NIH funding is the Stanford Chemistry Department's Mass Spectrometry laboratory. The scope of the new proposal has been broadened to include structural insights derived from a variety of sources in addition to mass spectral data, now including other spectroscopic, chemical, mechanistic and reaction data. This will make this research useful to other structure elucidation resources supported by NIH/BRP. Relating our research to analytical resources promotes an intimate working relationship between those responsible for program development and application and those collecting data on important new molecular structures.

During the current period we have addressed the problem of collaboration with a broader community by making our facilities available for the use of CONGEN to solve actual structure problems and to obtain critiques to find areas needing improvement (see question 3 below also). Although we have been able to assist many persons with structure elucidation problems (within the limits of our resources of personnel and equipment), we have had to refuse access to some. Broadening our scope introduces the possibility of a larger community of scientists interested in using our programs. We will maintain our current policy of assisting those we are able to, while working toward wider program availability (see question 4 below also).

2. What is the relationship to the SUMEX-AIM resource?

The goals of the SUMEX-AIM resource and the DENDRAL resource-related research project are quite distinct. The goals of SUMEX are 1) to develop a national community of projects applying "artificial intelligence" computer science techniques to medical research areas and 2) to investigate the utility of current computer communication technology for sharing computer resources and scientific exchange within the community. The DENDRAL project focuses on developing and applying AI techniques specifically to chemical structure elucidation problems in biochemistry. In this sense DENDRAL is a user of the SUMEX resource for the development of its programs, as a vehicle for making them available to collaborating medical researchers, and as an active participant in the SUMEX-AIM community. Data on DENDRAL use of the SUMEX-AIM resource are shown in Attachment A.

There is much communication between the SUMEX staff and DENDRAL staff, a realization of the hope under which SUMEX was formed. The DENDRAL project has developed all of its own programs on SUMEX. We have not asked SUMEX personnel to work on programs that are specific to DENDRAL, since this is not their charter. DENDRAL project personnel have made suggestions for improvement of SUMEX systems programs (as have many other user projects) such as text editors (SOS, TV), languages (INTERLISP, SAIL), manuscript preparation programs (PUB, SPELL), bulletin board programs (BBD, POST), and message handling programs (MSG, SNDMSG). These have been considered on a one by one basis by the SUMEX project and suggestions meeting overall community needs were scheduled for implementation and others not. SUMEX has provided assistance to DENDRAL, as it has to other user projects, for interfacing operational programs to the broader SUMEX user community.

Also, some of our research ideas have been developed by other SUMEX users. In particular, many of the production system ideas developed by DENDRAL people have been incorporated in the MYCIN program. We have discussed them extensively with Professor Todd Wipke at Santa Cruz, Professor Ken Colby at UCLA, and many others. We feel that this dissemination is exactly the intent of the SUMEX resource.

One of the major efforts of the SUMEX staff is the MAINSAIL project whose goal is to develop a subset of the SAIL language that will run on several different machines. We are certainly interested in the outcome of that and are willing to consult on that project when asked. We feel that DENDRAL programs will provide a nice test case for the MAINSAIL effort but there has been no mix in the personnel involved.

Some individuals, primarily support persons involved in maintenance of the facilities, spend time on both projects and in these cases the percent of time charged to DENDRAL is less than 100%. This arises from the needs of the projects, personal desires of the individuals, and our wish to make the best use of available talent. Data on personnel supported partly by both DENDRAL and SUMEX are shown in Attachment B.

3. What scientific problems have been solved by collaborators using DENDRAL programs? How has their use affected the development of DENDRAL programs?

The goal of DENDRAL is to develop a set of techniques that can be used by structure elucidation chemists throughout the nation. In pursuing that goal, we have found that suggestions from chemists with important biomedical structure elucidation problems are indispensable. We have actively sought new collaborators (through scientific papers presented at conferences, published articles, seminars, workshops, and personal contacts). The level of satisfaction of those users indicates our responsiveness to their suggestions as well as the level of performance of the programs. The names of the collaborators with indications of the problems they are working on and some indication of suggestions they have made to further our research are given in Attachment C.

These contacts with users solving real biochemical research problems have been essential to the continued development of the DENDRAL programs. Almost all developments over the past three years, and almost all new, proposed research, is in direct response to important applications. This is reflected in our recent publication record, which is oriented heavily to problems in chemistry, and the gradual shift in personnel to persons who are trained primarily as chemists but who also have computer science expertise, rather than the other way around. In our own group, chemists are responsible both for new applications and much of the actual algorithm design and program writing. Thus we have a situation where important new application areas strongly direct the course of program development.

Our current CONGEN program incorporates many features which were specifically implemented either to facilitate our own handling of interactions with other scientists (category A) or as a direct result of suggestions made by other persons (category B).

Category A: General DRAW programs to enable scientists with a variety of terminals to communicate structural information with the program or with us; GRIPE to register complaints and suggestions automatically; BUGOUT to save for our evaluation a complete copy of CONGEN in its current state when a suspected or actual program error has occurred; RECORD to record for our examination a complete interactive session with CONGEN so that we can assist scientists in more effective use of the program.

Category B: Incorporation of aromaticity and its analysis; extended features of EDITSTRUC including variable bond orders, variable aromatic types of atoms, specification of variable length chains of atoms; isoprene constraints; interrupt features to inform the scientist on the number of structures constructed and to allow him/her to examine them; and a preliminary estimator to report the progress of CONGEN toward completion, invaluable for estimating the time required for long, compute bound jobs. Current work on (1) extension of the "tagging" concept to substructures in general for efficient expression of constraints such as those from <sup>13</sup>CMR; (2) a new imbedder with extended capabilities for constraint testing for greater efficiency; and (3) extensions of reaction sequences and development of the MSPRUNE and MDGGEN programs for general analysis of mass spectra in the absence of class-specific fragmentation rules are all the result of using CONGEN in actual problems and finding areas for improvement.

#### 4. What new research is proposed? What has been done?

We have extracted specific items from the proposal that we feel constitute innovative research and have listed them here. For perspective, we also show many of our past accomplishments and our present work.

## A. New Research

The new research proposed is broken into categories and described in the proposal at the places indicated.

### (i) Structure Elucidation Programs:

Constraints Interpreter for CONGEN	(sec 3.1.1, p.15)
Help System for CONGEN	(sec 3.1.2, p.17)
Experiment Planning Program	(sec 3.2.1, p.18)
Reaction Chemistry Program(s)	(sec 3.2.2, p.21)
Class-Independent M.S. Interpreter	(sec 3.2.3, p.25)
C13 NMR Interpreter	(sec 3.2.4, p.29)

### (ii) Theory Formation Programs:

Feedback Loops	(sec 3.3.1.1, p.30)
Alternative Rule Models	(sec 3.3.1.2, p.31)
C13 NMR Rule Formation	(sec 3.3.2, p.32)
Generalization of Programs	(sec 3.3.3, p.34)

### (iii) New Applications:

Marine Natural Products	(sec 3.4.1, p.35)
Analysis of Body Fluids	(sec 3.4.2, p.36)
Cyclization Products	(sec 3.4.3, p.39)
Steroids (C13 NMR)	(sec 3.4.4, p.42)

### (iv) Data Collection and Data Reduction:

Increased Sensitivity	(sec 3.6.1, p.47)
-----------------------	-------------------

## B. Past Research Highlights

Some of the major accomplishments and developments are given here for each of several lines of research.

## INSTRUMENTATION

- 1970 Existing MS-9 interfaced to ACME computer system for acquisition and reduction of high resolution mass spectrometric data.
- 1970-71 First version of HRMS data reduction programs developed, tested and applied.
- 1972 Varian-MAT 711 high resolution mass spectrometer delivered.
- 1972-73 711 interfaced to ACME and existing programs.
- 1973 Introduction of data reduction concepts based on computed measures of instrument performance.
- 1973-74 First Version of program for data reduction from combined gas chromatograph/high resolution mass spectrometer system.

- 1974 PDP 11/45 purchased and delivered to supplant ACME.
- 1974-75 Reprogramming of high resolution data reduction software. Addition of instrument evaluation displays.
- 1975-76 Improvements in data reduction software; introduction of doublet resolver.
- 1975-76 Applications of GC/HRMS system to chemical problems.
- 1976 Pending proposal for development of higher sensitivity GC/HRMS system.

#### STRUCTURE GENERATOR

- 1962-64 Lederberg conceives algorithm for acyclic structure generation.
- 1962-64 Concept generalized to ring systems based on vertex-graphs.
- 1965 First version of acyclic structure generator program.
- 1966 Concept of BADLIST introduced.
- 1967 Concept of superatoms and GOODLIST introduced.
- 1968-69 Program for acyclic structure generation with constraints developed and improved.
- 1969-71 Application of acyclic structure generator to interpretation of mass spectra of acyclic compounds. Mass spectral peaks and rules and, later, NMR data used as constraints.
- 1970 Extension of acyclic generator to monocyclic molecules based on pairwise removal of hydrogen atoms. Approach limited to monocyclics and prone to duplication.
- 1971-72 Algorithm devised for construction of ring systems based on vertex-graphs.
- 1972-73 PLANNER written with an ad hoc generator of cyclic structures.
- 1972-73 Algorithm for complete structure generation devised.
- 1972-73 First program for generation of all isomers of a given empirical formula written; duplicate structures avoided prospectively.
- 1973-74 Superatom representation and imbedder developed.
- 1974 First user interface and constraints mechanism added to structure generator.
- 1974-75 First version of CONGEN developed for export via SUMEX.
- 1975-76 Application of CONGEN to chemical problems begun. Major improvements to user interface and constraints implementation. Depth-first version introduced.

- 1976 Introduction of mass spectrometry functions to CONGEN. First version of reaction sequences implemented.
- 1976 Pending proposal for development of and extensions to CONGEN as a general tool for computer-assisted structure elucidation.

THEORY FORMATION  
(Meta-DENDRAL)

- 1971 Prototype version of INTSUM running.
- 1972-73 INTSUM used to analyze mass spectra of estrogenic steroids.
- 1973-74 Prototype version of RULEGEN running.  
INTSUM extended.  
Prototype version of RULEMOD running.
- 1975 INTSUM used to analyze mass spectra of capnellanes, pregnanes, ketoandrostanes.
- 1975-76 INTSUM-RULEGEN-RULEMOD used to analyze mass spectra of ketoandrostanes. Also used on aromatic acid spectra and marine sterol spectra.
- 1976 Preliminary modifications of programs to infer C13 NMR rules from spectra. Applications to aliphatic hydrocarbons and amines.

DENDRAL PLANNER

- 1967-68 First version of Preliminary Inference Maker.
- 1968-69 Application to aliphatic ketone alcohols, ethers, amines (with generator).
- 1969-70 Preliminary Inference Maker extended to handle all saturated aliphatic monofunctional compounds. Introduction of NMR inferences.
- 1970-71 Application to thioethers, thiols, as well as alcohols, ethers and amines.
- 1972-73 Reconceptualization of planning program to generalize notion of planning rule generation. Introduction of cyclic structures. Introduction of feedback loops.
- 1973 Application to estrogenic steroids.
- 1973-74 C13 NMR Planner written. Application to aliphatic amines.
- 1974-75 MOLION developed to infer molecular ions of unknowns. Incorporated in DENDRAL (MS) PLANNER.

Attachment C. Problems of research collaborators related to structure elucidation solved with the aid of DENDRAL programs.

1. Dr. Roger Hahn, Syracuse University. While at Stanford he used CONGEN to help solve the structures of photoproducts by obtaining all possibilities under available constraints and designing NMR experiments to differentiate the possibilities. This work will be published soon.
2. Dr. William Epstein, University of Utah. During a demonstration of CONGEN, he posed a problem to verify that the structural possibilities he determined for an unknown were in fact all possibilities. The structure of methyl santoninate has been published (see Epstein, et al., J.C.S. Chem. Commun., 590 (1975)).
3. Drs. William Milne and Henry Fales, DCRT/NIH. They have used CONGEN and MOLION through an EXO-DENDRAL account. One problem involved questions on benzene isomers, another on aspects of the (known) structure of a quinolinone, for which CONGEN and MOLION were used. We have responded to questions on use of the programs, but have not been in close contact on their specific applications.
4. Dr. Richard Feldmann, DCRT/NIH. We have worked closely with him on problems of graphical structure display, on both graphics and hard copy terminals.
5. Dr. M. J. Goldstein, Cornell University. Using CONGEN, we have provided him with structural possibilities for a new, tricyclic C<sub>11</sub>H<sub>12</sub> hydrocarbon. There were many constraints on this problem, all of which were incorporated in arriving at 92 possibilities, of which all but one were eliminated by proton NMR experiments. This work is to be published.
6. Dr. F. W. McLafferty, Cornell University. During a demonstration of CONGEN we provided him with some sets of isomers of ion structures in support of his work in structures of gaseous ions.
7. Dr. Todd Wipke, University of California, Santa Cruz. We are working closely with him on problems of symmetry in structure manipulation, problems which are critical to the efficient use of programs such as CONGEN and the proposed extensions to reaction sequences.
8. Dr. Clair Cheer, University of Rhode Island. While on sabbatical at Stanford, Dr. Cheer has worked on a number of structure elucidation problems using CONGEN including Briareine D (see p. 33 of 1976 annual report) and [±]-Palustrol (see Cheer et al., Tetrahedron Letters, 1807 (1976), attached). Work is continuing on the structure of another marine natural product, presumably a cembrenolide, for which there are currently seven possibilities.



9. Dr. Jerrold Karliner, Ciba-Geigy Corporation. Dr. Karliner has solved several structural problems using CONGEN, including material with flame retardent properties, an impurity in a production sample and nitrogen heterocycles being investigated for pharmacological activity. CONGEN enabled reduction of the number of possibilities to the point where subsequent experiments led to unambiguous structural assignment (see attached letter).

10. Dr. Gino Marco, Ciba-Geigy Corporation. He has used CONGEN to help solve structures of conjugates of pesticides with sugars and amino acids.

11. Dr. Milton Levenberg, Abbott Laboratories. He has worked on the structure of a compound with mild antibiotic activity, isolated from a fermentation broth. There are currently ten structural possibilities, reduced to that number from the 33 initially determined using CONGEN by additional experimental data.

12. Dr. David Pensak, DuPont. He is currently learning to use CONGEN and plans to evaluate its utility for structural problems of some of his coworkers.

13. Dr. Douglas Dorman, Eli-Lilly. He is using CONGEN to assist in structure elucidation of metabolites of microorganisms shown to have pharmacological activity. He has worked on five such problems, including a current one where the developing MSPRUNE capabilities are being used (see attached letter).

14. Dr. L. Minale, Napoli, Italy. We have worked with him by sending him structural alternatives for proposed structures for some marine natural products (Pallescensins, Tetrahedron Letters, 1417 (1975)) and cyclic diethers from the lipid fraction of a thermophilic bacterium (J. C. S. Chem. Commun., 543 (1974)) (see attached letter).

15. Dr. K. Nakanishi, Columbia University. We have worked with him by sending him structural possibilities for termite defense compounds (structure finally solved by X-ray crystallography) This trial (see attached correspondence for the flavor of the collaboration) plus a live demonstration to one of his students has resulted in efforts toward continued collaboration on other insect defense secretions and exploration of the possibility of his direct access to SUMEX.

16. Dr. L. Dunham, Zoecon Corporation. We have collaborated with him on the use of INTSUM for mass spectral fragmentation studies of insect juvenile hormones (see ref. 67 in annual report).

17. Dr. A. G. Gonzales, Tenerife, Spain. We have recently sent him structural alternatives for constituents of Laurencia Perforata (Tetrahedron Letters, 2499 (1975)), and expect to continue discussions on the structures of these compounds (see attached letter).

18. Dr. T. Irie, Sapporo Japan. We have recently sent him structural alternatives to published structures on constituents of Laurencia Glandulifera (Tetrahedron Letters, 821 (1974)) and expect to continue discussions on this problem. (see attached letter to Prof. Irie).

19. Dr. C. J. Persoons, Delft. We have corresponded with him on structural alternatives for cockroach sex pheromones (Periplanone-B (Tetrahedron Letters, 2055 (1976))), and he has agreed to further collaboration on new problems.

20. Dr. F. Schmitz, University of Oklahoma. We explored for him structural alternatives for an unknown diterpenoid hydrocarbon. We obtained 25 possibilities, of which only four obeyed the isoprene rule.

21. Dr. J. Baker, Roche Institute of Marine Pharmacology, Australia. We plan collaboration with Dr. Baker on the sterol fractions of various marine organisms and are exploring ways for him to access CONGEN.

22. Dr. E. VanTamelon, Stanford University. We have used the developing reaction features of CONGEN to explore structural possibilities for both chemical and biogenetic cyclization products of squalene-oxide congeners. We have suggested alternatives to proposed structures and helped to design experiments to differentiate them.

23. Dr. J. C. Braekman, Brussels. Dr. Braekman visited Stanford as a part of continuing collaboration in marine chemistry with Dr. Tursch's group. While at Stanford he explored use of CONGEN for use in current problems in marine natural products, and worked on the problems of Drs. Irie and Gonzales (see above). He is currently exploring access to CONGEN from Brussels, via ARPANET (see attached letter).

#### 24. DENDRAL Group Applications

Because most of our applications are described in detail in our publications, we will only itemize them here with pointers to the reference list at the end of the 1976 annual report (an Appendix to the proposal).

- a) Isomers of chlorinated hydrocarbons (ref. 46).
- b) Structural and positional isomers of ring systems (ref. 47).
- c) Ion structures; structures of ions from decomposition of triethylamine molecular ions (ref. 55).
- d) Structural isomerism of terpenoid systems (ref. 60).
- e) Fragmentations of ketoandrostane molecular ions (ref. 58).
- f) Reaction sequences applied to isomerization and cyclization and rearrangement processes (ref. 59 and also #22, above).
- g) Currently we are focussing our attention on computer-based approaches to analysis of marine sterols, chemical fractions of body fluids and <sup>13</sup>CMR rule formation and prediction. These application areas are described in detail in the annual report and proposal.