DR. LEDERBERG

RESOURCE-RELATED RESEARCH
COMPUTERS AND CHEMISTRY
(RR-00612 RENEWAL APPLICATION)

Submitted to the

BIOTECHNOLOGY RESOURCES BRANCH

OF THE

NATIONAL INSTITUTES OF HEALTH

December, 1973

School of Medicine
Stanford University

SECTION I

Form Approved
Budget Bureau No. 68-R0249

DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE
PUBLIC HEALTH SERVICE

GRANT APPLICATION

LEAVE BLANK		
TYPE	PROGRAM	NUMBER
REVIEW GROUP		FORMERLY
COUNCIL *(Month, Year)*		DATE RECEIVED

TO BE COMPLETED BY PRINCIPAL INVESTIGATOR *(Items 1 through 7 and 15A)*

1. TITLE OF PROPOSAL *(Do not exceed 53 typewriter spaces)*

Resource Related Research - Computers and Chemistry (RR-00612 renewal)

2. PRINCIPAL INVESTIGATOR

2A. NAME *(Last, First, Initial)*

Djerassi, Carl

2B. TITLE OF POSITION

Professor of Chemistry

2C. MAILING ADDRESS *(Street, City, State, Zip Code)*

Department of Chemistry
Stanford University
Stanford, California 94305

2D. DEGREE

Ph.D.

2E. SOCIAL SECURITY NO.

[redacted]

2F. TELEPHONE DATA — Area Code: 415 — TELEPHONE NUMBER AND EXTENSION: 321-2300, Ext. 2783

2G. DEPARTMENT, SERVICE, LABORATORY OR EQUIVALENT *(See Instructions)*

Department of Chemistry

2H. MAJOR SUBDIVISION *(See Instructions)*

School of Humanities and Sciences

3. DATES OF ENTIRE PROPOSED PROJECT PERIOD *(This application)*

FROM: 5/1/74 THROUGH: 4/31/79

4. TOTAL DIRECT COSTS REQUESTED FOR PERIOD IN ITEM 3

$1,350,795.00

5. DIRECT COSTS REQUESTED FOR FIRST 12-MONTH PERIOD

$276,197.00

6. PERFORMANCE SITE(S) *(See Instructions)*

Department of Genetics,
Department of Chemistry, and
Department of Computer Science
Stanford University

7. Research Involving Human Subjects *(See Instructions)*

A. ☐ NO B. ☒ YES Approved: _____ Date

C. ☒ YES – Pending Review

8. Inventions *(Renewal Applicants Only - See Instructions)*

A. ☒ NO B. ☐ YES – Not previously reported

C. ☐ YES – Previously reported

TO BE COMPLETED BY RESPONSIBLE ADMINISTRATIVE AUTHORITY *(Items 8 through 13 and 15B)*

9. APPLICANT ORGANIZATION(S) *(See Instructions)*

Stanford University
Stanford, California 94305
IRS No. 94-1156365
Congressional District No. 17

10. NAME, TITLE, AND TELEPHONE NUMBER OF OFFICIAL(S) SIGNING FOR APPLICANT ORGANIZATION(S)

c/o Sponsored Projects Office
Telephone Number (s) (415) 321-2300, X2883

11. TYPE OF ORGANIZATION *(Check applicable item)*

☐ FEDERAL ☐ STATE ☐ LOCAL ☒ OTHER *(Specify)*

Private, non-profit University

12. NAME, TITLE, ADDRESS, AND TELEPHONE NUMBER OF OFFICIAL IN BUSINESS OFFICE WHO SHOULD ALSO BE NOTIFIED IF AN AWARD IS MADE

K. D. Creighton
Deputy Vice Pres. for Business & Finance
Stanford University
Stanford, California 94305

Telephone Number (415)321-2300, X2551

13. IDENTIFY ORGANIZATIONAL COMPONENT TO RECEIVE CREDIT FOR INSTITUTIONAL GRANT PURPOSES *(See Instructions)*

20 School of Humanities and Sciences

14. ENTITY NUMBER (Formerly PHS Account Number)

458210

15. CERTIFICATION AND ACCEPTANCE. We, the undersigned, certify that the statements herein are true and complete to the best of our knowledge and accept, as to any grant awarded, the obligation to comply with Public Health Service terms and conditions in effect at the time of the award.

SIGNATURES (Signatures required on original copy only. *Use ink, "Per" signatures not acceptable)*	A. SIGNATURE OF PERSON NAMED IN ITEM 2A	DATE
	B. SIGNATURE(S) OF PERSON(S) NAMED IN ITEM 10	DATE

NIH 398 (FORMERLY PHS 398
Rev. 1/73

SECTION 1

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
PUBLIC HEALTH SERVICE

RESEARCH OBJECTIVES

LEAVE BLANK

PROJECT NUMBER

NAME AND ADDRESS OF APPLICANT ORGANIZATION

Stanford University
Stanford, California 94305

NAME, SOCIAL SECURITY NUMBER, OFFICIAL TITLE, AND DEPARTMENT OF ALL PROFESSIONAL PERSONNEL ENGAGED ON PROJECT. BEGINNING WITH PRINCIPAL INVESTIGATOR

Carl Djerassi [redacted] Professor of Chemistry, Department of Chemistry; Joshua Lederberg, [redacted] Professor of Genetics, Department of Genetics; Edward Feigenbaum [redacted] rofessor of Computer Science, Department of Computer Science; Bruce Buchanan, [redacted] Research Computer Scientist, Department of Computer Science; Alan Duffield, [redacted], Research Associate, Department of Genetics; Dennis Smith, [redacted], Research Associate, Department of Genetics; Natesa Sridharan, [redacted], Research Associate, Department of Computer Science; Harold Brown [redacted], Research Associate, Department of Computer Science; Geoff Dromey, SS# applied for-to be supplied at a later date, Department of Computer Science.

TITLE OF PROJECT

Resource-Related Research -- Computer and Chemistry

USE THIS SPACE TO ABSTRACT YOUR PROPOSED RESEARCH. OUTLINE OBJECTIVES AND METHODS. UNDERSCORE THE KEY WORDS (NOT TO EXCEED 10) IN YOUR ABSTRACT.

The objectives of this research program are the development of innovative computer and biochemical analysis techniques for application in medical research and closely related aspects of investigative patient care. We will apply the unique analytical capabilities of gas chromatography/mass spectrometry (GC/MS) with the assistance of data interpreting computer programs utilizing artificial intelligence techniques, to investigate the chemical constituents of human body fluids in a variety of clinical contexts. Specific subtasks of this program include; 1) the application of artificial intelligence (AI) techniques to programs capable of interpreting mass spectra from basic principles as well as extending mass spectral theory by analysis of solved spectrum-structure examples, 2) the extension of GC/MS data systems to provide stand-alone capabilities for collecting low and high resolution mass spectral and metastable ion data, 3) the application of GC/MS and AI techniques to analysis of biomolecular structure elucidation problems of a large number of collaborators, and 4) the extension of artificial intelligence techniques to an interactive system for computer assisted structure elucidation based on a variety of data.

LEAVE BLANK

SECTION II – PRIVILEGED COMMUNICATION

DETAILED BUDGET FOR FIRST 12-MONTH PERIOD

FROM 5/1/74 THROUGH 4/31/75

DESCRIPTION (Itemize)		TIME OR EFFORT %/HRS.	AMOUNT REQUESTED (Omit cents)		
PERSONNEL					
NAME	TITLE OF POSITION		SALARY	FRINGE BENEFITS	TOTAL
	PRINCIPAL INVESTIGATOR				
DETAILED SALARY DATA LISTED SEPARATELY ON ATTACHED SHEET					
			163,935	28,962	192,897
CONSULTANT COSTS					-0-
EQUIPMENT					
Equipment Purchase (First Year Items Only):					
DEC GT-40 Display Terminal					13,400
PDP 11/20 Upgrade					34,000
Equipment Maintenance:					
PDP-11 (DEC Contract)					4,200
MAT-711 (Parts, etc.)					6,500
SUPPLIES Electronics Supplies					4,400
GC Supplies (Columns, gases, etc.)					1,000
Liquid Nitrogen					1,000
Chemicals, glassware, stock, etc.					1,500
Data Recording Media					1,000
Minicomputer Supplies					700
TRAVEL DOMESTIC					1,200
TRAVEL FOREIGN					-0-
PATIENT COSTS (See instructions)					-0-
ALTERATIONS AND RENOVATIONS					-0-
OTHER EXPENSES (Itemize)					
Publications, telephone, office supplies, postage					4,000
Computer Terminal Lease (4)					5,400
Computer Usage - 370/158 (First Year Item Only)					5,000
TOTAL DIRECT COST (Enter on Page 1, Item 5) →					276,197

INDIRECT COST (See Instructions)

DATE OF DHEW AGREEMENT:

_____ % S&W*

47 % ~~TDC~~ NTDC - June 26, 1973

☐ WAIVED

☐ UNDER NEGOTIATION WITH: _____

*IF THIS IS A SPECIAL RATE (e.g. off-site), SO INDICATE.

DETAILED SALARY DATA
NIH GRANT #RR-00612
5/1/74-4/31/75

	% Effort	Salary	Fringe Benefits	Total
PRINCIPAL INVESTIGATORS:				
C. Djerassi	10	-0-	-0-	-0-
J. Lederberg	10	-0-	-0-	-0-
E. Feigenbaum	10	2,910	514	3,424
RESEARCH ASSOCIATES:				
B. Buchanan (1)	50	7,000	1,237	8,237
A. Duffield	25	6,195	1,094	7,289
D. Smith	100	16,200	2,862	19,062
N. Sridharan	100	16,050	2,835	18,885
H. Brown	100	16,200	2,862	19,062
G. Dromey	100	15,500	2,738	18,238
PROGRAMMERS:				
W. White	100	14,400	2,545	16,945
R. Tucker	100	14,100	2,491	16,591
SENIOR RESEARCH ASSISTANT:				
A. Wegmann	100	15,000	2,650	17,650
ELECTRONICS ENGINEER:				
N. Veizades	60	11,670	2,062	13,732
GLASS BLOWER/MACHINEST:				
E. Steed	25	4,410	779	5,189
RESEARCH ASSISTANTS:				
L. Masinter	100	5,070	895	5,965
M. Stefik	100	4,915	868	5,783
To Be Appointed	100	4,915	868	5,783
SECRETARIAL SUPPORT:				
K. Wharton	100	9,400	1,662	11,062
TOTAL:		$163,935	$28,962	$192,897

(1)Dr. Buchanan's salary charges do not begin until 9/1/74 at which time his NIH Research Career Development Award expires.

BUDGET ESTIMATES FOR ALL YEARS OF SUPPORT REQUESTED FROM PUBLIC HEALTH SERVICE
DIRECT COSTS ONLY (Omit Cents)

DESCRIPTION		1ST PERIOD (SAME AS DETAILED B)	ADDITIONAL YEARS SUPPORT REQUESTED *(This application only)*					
			2ND YEAR	3RD YEAR	4TH YEAR	5TH YEAR	~~XXXXXX~~ TOTAL	7TH YEAR
PERSONNEL COSTS		192,897	210,611	225,129	240,630	257,383	1,126,650	
CONSULTANT COSTS *(Include fees, travel, etc.)*		-0-	-0-	-0-	-0-	-0-	-0-	
EQUIPMENT		58,100	11,770	12,947	14,241	15,665	112,723	
SUPPLIES		9,600	6,920	7,612	8,370	9,207	41,709	
TRAVEL	DOMESTIC	1,200	1,320	1,452	1,597	1,757	7,326	
	FOREIGN	-0-	-0-	-0-	-0-	-0-	-0-	
PATIENT COSTS		-0-	-0-	-0-	-0-	-0-	-0-	
ALTERATIONS AND RENOVATIONS		-0-	-0-	-0-	-0-	-0-	-0-	
OTHER EXPENSES		14,400	10,340	11,374	12,511	13,762	62,387	
TOTAL DIRECT COSTS		276,197	240,961	258,514	277,349	297,774	1,350,795	

TOTAL FOR ENTIRE PROPOSED PROJECT PERIOD *(Enter on Page 1, Item 4)* ⟶ $ 1,350,795

REMARKS: *Justify all costs for the first year for which the need may not be obvious. For future years, justify equipment costs, as well as any significant increases in any other category. If a recurring annual increase in personnel costs is requested, give percentage. (Use continuation page if needed.)*

See following pages for budget justification.

Budget Justification

The availability of existing equipment - including the mass spectrometer and SUMEX computer - avoids the need for requesting funds for major laboratory items and substantial computing costs. Thus, the major expense in the resulting budget is for personnel. We feel that the personnel listed here are necessary to carry out the research, as justified below. Recurring costs are about $227,000 per year. First year expenditures are higher to provide the instrumentation necessary for mass spectrometry service in the first year.

We are requesting funds for five years to coincide with the funding of the AIM-SUMEX resource, to which we hope to make significant constributions.

This budget overlaps slightly with the budget for the Genetics Research Center (J. Lederberg, Principal Investigator). Dr. Alan Duffield's 25% salary budgeted here is covered by the other budget (where 100% of his salary is budgeted). 10% of Ms. Annemarie Wegmann's salary is covered there (with 100% of her salary budgeted here). These are the only overlapping items. We have no official notification of Genetics Center funding; if the present proposal is successful, the Genetics Center budget will be adjusted accordingly.

In the five-year budget, salaries are increased by 6% per year and staff benefits are computed at 17% for the period 5/74-8/74, 18% for the period 9/74-8/75, and are increased 1% per year thereafter, based on current University projections. Other budget categories are increased by 10% per year to account for inflation.

Personnel:

BRUCE G. BUCHANAN
Dr. Bruce Buchanan holds an NIH Research Career Development Award to work on applications of artificial intelligence to health-related problems, including theory formation by computer. His work on those aspects of this grant is thus consistent with the Development Award. Half-time support is requested after the third year of the Development Award (starting September, 1974) to cover the contingency that the award will not be extended to the full five years. These funds will be returned if the Award is extended.

DENNIS H. SMITH
Dr. Dennis H. Smith has been a member of the DENDRAL project since July, 1971. He has been responsible for the MS and its computer support, and has been involved in the application of the AI programs to structural studies of biomedically important compounds, primarily steroids. These responsibilities will continue in the future, with particular emphasis on providing the mass spectrometer/AI program link to the user community and its mass spectrometry and general structure elucidation needs, and in providing the necessary chemical knowledge and input for development of the computer programs and user interfaces for the proposed computer assisted structure elucidation effort.

ALAN DUFFIELD
Dr. Alan Duffield is the senior scientist in charge of the mass spectrometry facilities of the GRC. Because of his expertise in the analysis of mass spectra from various fractions of human body fluids, he will provide the link between the structure elucidation techniques of this proposal and other scientists with similar problems. The GC/HRMS facilities are also expected to provide service to the Genetics Center for high resolution analysis of compounds isolated from body fluids.

NATESA SRIDHARAN
Dr. Sridharan will be responsible for developing interface routines that allow new researchers to make use of the structure elucidation programs. We expect these routines to accept information about a research problem, in semi-formal terms, and translate it into a format the program can use. They should be complete enough so that individual researchers do not need to know about the inner workings of the programs. In addition, he will continue to help Dr. Brown and Mr. Masinter with development of the cyclic generator program. (Within a few days of this writing, Dr. Sridharan has decided to take a leave of absence. During his absence we will recruit another Research Associate to perform his duties.)

HAROLD BROWN
Dr. Harold Brown's knowledge of graph theory and combinatorial mathematics is essential to the development of the cyclic structure generator. Many problems with development and implementation of this program have required sophisticated, new mathematical solutions worked out by Dr. Brown. For example, generating the dictionary of cyclic graphs and assembling substructures involve problems in graph theory that Dr. Brown is currently working on.

Dr. Brown has submitted a proposal to the NSF to cover his salary for this research. If that grant is awarded, funds requested here for his salary will not be needed.

R. GEOFF DROMEY
Dr. Geoff Dromey is a chemist with strong computer science interests who has been associated with the project since September, 1973. He has become familiar with many aspects of the DENDRAL performance programs and will be expected to help outside researchers use those programs. In addition, he will be developing new programs, such as the program for molecular ion determination from mass spectra.

WILLIAM C. WHITE
Mr. William White provides high-level programming support for the theory formation programs, including helping to devise new programs in response to new research problems as well as implementing them. He wrote almost all of the LISP code for the INTSUM program, for example, and is currently responsible for the RULEGEN program.

MS. ANNEMARIE WEGMANN
Ms. Annemarie Wegmann is the Senior Research Assistant in charge of the GC/HRMS system. She was formerly head of Hewlett-Packard's Palo Alto gas chromatography applications laboratory and has been responsible for the operation of the GC/MS system since the

delivery to our laboratory of the MAT-711 (November, 1971). Her technical ability is absolutely essential to the continued operation and development of the mass spectrometry facility.

INSTRUMENT SUPPORT PERSONNEL

Messers. Veizades and Steed will assist part time in maintaining the GC/MS system. Mr. Veizades is an Electronics Engineer who is responsible for the electronic and mechanical systems as well as providing the necessary voltage read-out and control development for the metastable analysis data system. Mr. Steed is a Research Engineer responsible for the system glasswork and vacuum system maintenance.

ROBERT TUCKER
Mr. Robert Tucker implements and maintains the computer programs for data acquisition and reduction of MS data. This includes translating existing PL/ACME into FORTRAN and PDP-11 assembly language. In addition, he will be responsible for improving these programs for repetitive HRMS scans, implementing the multiplet resolution algorithm and the software necessary for semi-automated collection of metastable ion data.

LARRY M. MASINTER
Mr. Larry Masinter, Research Assistant, will continue to work with Drs. Lederberg and Brown on the development of the cyclic structure generator. His LISP expertise has been an invaluable resource for every member of the research team.

MARK STEFIK
Mr. Mark Stefik, Research Assistant, combines two years of experience on the ACME/MS data acquisition system with a long-term commitment to computer science. He has developed interactive library search capabilities for the mass spectrometer and will continue to improve them. His knowledge of the data acquisition computer programs will be very valuable in assisting initial translation of those programs into FORTRAN (from PL/ACME code) for the extended PDP-11/20 system.

RESEARCH ASSISTANT - unnamed
We have interviewed two prospective Research Assistants, both of whom have broad chemical experience and strong computer science interests. We request funds to hire one of them to provide additional links between computer science techniques and structure elucidation problems.

SECRETARIAL SUPPORT
One full-time secretary is necessary for the secretarial support of this number of scientists. Ms. Kathleen Wharton is now with the Computer Science group.

EQUIPMENT PURCHASE:
As discussed in the text (Section III.A), in the first year we plan to augment our existing PDP-11/20 computer (4k memory) to allow its operation as a stand-alone data system. We plan to add 16k of memory ($3,000), a floating print arithmetic unit ($7,500), an industry compatible tape drive ($9,000), a disk drive (10,500), a low speed communications interface ($1,000), and a bootstrap

loader and clock ($1,200). These devices together with state sales tax total to $34,000. The prices quoted are representations of the most cost-effective suppliers of the respective devices we have been able to locate. We will continue to review the market before implementation to maximize technical and cost performance.

As stated above, we plan to provide interface programs to provide the communication link between the users and the programs. The universal language of molecular structure is diagrammatic representation of the structures, drawn usually in two dimensions (or as two-dimensional representations of three dimensional information). Therefore, we feel that a graphics terminal such as the DEC GT-40 is necessary for effective sharing of the programs among Stanford users. The GT-40 terminal is a good choice for performing this structural display task, for a number of reasons. Programs are available for input and output of structural information which can be modified to run on a GT-40 (e.g., we have just implemented on an experimental basis routines made available to us by R. Feldman, NIH); Sophisticated structural display programs have been written especially for a GT-40 which we would hope to mount; and the AIM-SUMEX resource will specifically support one GT-40 for use by the SUMEX staff. This terminal will be physically located in the MS laboratory since all of the users will interact with that laboratory.

EQUIPMENT MAINTENANCE:
Maintenance is budgeted for the proposed stand-alone PDP-11/20 system under DEC contract based on current prices. Also included is a budget for maintenance of the MAT-711 system. This estimate is based on our experience with parts replacements to date. We will provide the necessary maintenance manpower (see personnel justification) because Varian cannot provide adequate service.

SUPPLIES:
Supplies are budgeted in various categories based on our operating experience to date. Electronics supplies include parts necessary for maintaining our electronics and test equipment ($1,000) as well as parts in the first year for the metastable ion data system ($3,200). These comprise several D/A and A/D converters for accelerating voltage, ESA voltage, and magnetic field control as well as parts to upgrade the Hall probe mass marker. GC supplies include carrier gases, columns, phases, syringes, septa, etc., for GC/MS operation. The liquid nitrogen is required for cold trap operation on the MAT-711. Chemicals, glassware, etc., include the various organic chemicals, glassware, apparatus, glass tubing, etc., needed to support the GC/MS laboratory operation. Data recording media include special uv sensitive recording paper for the MAT-711, paper for GC and instrumentation recorder, and calcomp paper and pens for ion currennt and spectrum plotting. Mini-computer supplies include paper, magnetic tape, ribbons, spare disk cartridges, etc., for data system operation.

TRAVEL:
The travel budget covers estimated needs (2 east coast and 2 west coast) trips for attending related professional meetings and interfacing potential program users nationally. Domestic travel is budgeted for two East Coast trips and two California trips per year among all personnel. No foreign travel is budgeted.

OTHER:

The "Other" budget includes operating telephone, office supplies, postage, reproduction, etc., support necessary for this project based on our previous experience. The "computer usage" allocation provides a continued limited usage of the 370/158 computer during the augmentation of the PDP-11/20 system. This cost does not appear in later years. Terminal rental covers four terminals to be distributed among the MS laboratory, the Computer Science Dept., and J. Lederberg's laboratory.

BIOGRAPHICAL SKETCHES

BIOGRAPHICAL SKETCH

(Give the following information for all professional personnel listed on page 3, beginning with the Principal Investigator. Use continuation pages and follow the same general format for each person.)

NAME	TITLE	BIRTHDATE *(Mo., Day, Yr.)*
Carl DJERASSI	Professor of Chemistry	October 29, 1923
PLACE OF BIRTH *(City, State, Country)*	**PRESENT NATIONALITY** *(If non-U.S. citizen, indicate kind of visa and expiration date)*	**SEX**
Vienna, Austria	U.S.A.	☒ Male ☐ Female

EDUCATION *(Begin with baccalaureate training and include postdoctoral)*

INSTITUTION AND LOCATION	DEGREE	YEAR CONFERRED	SCIENTIFIC FIELD
Kenyon College	A.B. (summa cum laude)	1942	Chemistry, Biology
University of Wisconsin	Ph.D.	1945	Organic chemistry, Biochemistry (minor)

HONORS Hon. D.Sc., Natl. Univ. of Mexico (1953), Kenyon College (1958), Worcester Polytechnic Institute (1972); Hon. Prof., Fed. Univ. Rio de Janeiro (1969). Member U.S. National Academy of Sciences, American Academy of Arts and Sciences, foreign member, Royal Swedish Academy of Sciences, German Academy of Natural Scientists (Leopoldina), Brazilian Academy of Sciences, (cont. below)

MAJOR RESEARCH INTEREST Nat. prod. chemistry (steroids, alkaloids, terpenoids, antibiotics) and chem. applications of physical methods (mass spec., optical rotatory dispersion, circular dichroism).

ROLE IN PROPOSED PROJECT: Principal Investigator

RESEARCH SUPPORT *(See instructions)*

Grant	Title	Period	Current Year	Total Budgeted	% Time Effort
NIH AM 04257	Mass Spectrometry in Organic and Biochemistry	10/1/70 to 9/30/75	$52,306	$316,016	10%
NIH GM AM 06840-15	Marine Chemistry with special emphasis on steroids	1/1/73 to 12/31/77	$75,650	578,180	18%

NSF Pending Grant Application #P3P3689, Magnetic Circular Dichroism in Organic Molecules, in the amount of $27,640.

RESEARCH AND/OR PROFESSIONAL EXPERIENCE *(Starting with present position, list training and experience relevant to area of project. List all or most representative publications. Do not exceed 3 pages for each individual.)*

Academic Experience:
Professor of Chemistry, Stanford University, 1959-present.
Associate Professor (1952-1954) and Professor (1954-1959), Wayne State University.

Industrial Research Experience:
Ciba Pharmaceutical Co., Summit, N.J.: Research Chemist, 1942-1943 and 1945-1949.
Syntex Corporation: Associate Director of Chemical Research (Mexico City) 1949-1952, Research Vice President (Mexico City) 1957-1960; (Palo Alto, California) 1960-1968, President, Syntex Research 1968-~~present~~ 1972, Zoecon Corporation (Palo Alto), President, 1968-~~present~~ 1972.

Editorial Boards:
(Current) Journal of the American Chemical Society, Steroids, Tetrahedron, Organic Mass Spectrometry.

(continued on next page)

Honors (cont.)
Mexican Academy for Scientific Investigation. Hon. Fellow of Phi Lambda Upsilon. Amer. Academy of Pharmaceutical Sciences, British Chemical Society and Mexican Chemical Society, Phi Beta Kappa. Numerous hon. lectureships including 1964 Centenary Lecturer (The British Chemical Society) and 1969 Annual Chemistry Lecturer, Royal Swedish Academy of Engineering. American Chemical Society Award in Pure Chemistry (1958), Baekeland Medal (1959), Fritzsche Award (1960). Intra-Science Research Foundation Award (1969). Freedman Patent Award of American Institute of Chemists (1971). Foreign Member, Royal Swedish Academy of Sciences (1972). D.Sc. (hon.), Worcester Poiytechnic Institute (1972). Scheele-Lecturer, Pharmaceutical Society of Sweden (1972); American Chemical Society's Award for Creative Invention (1973), National Medal of Science (1973).

U. S. GOVERNMENT PRINTING OFFICE : 1971 O - 451-736

11

RESEARCH AND/OR PROFESSIONAL EXPERIENCE (cont.)

Miscellaneous:
Chairman of the AAAS Gordon Research Conferences on Steroids and Natural Products (1952-1954); Member of American Pugwash Committee (1968 to present); Chairman of Latin America Science Board of National Academy of Sciences (1966-1968); Chairman of National Academy's Board on Science and Technology for International Development.

PUBLICATIONS

Author or co-author of six books and approximately 800 publications dealing with natural products (notably steriods, terpenoids, alkaloids and antibiotics), medicinal chemistry (primarily antihistamines, oral contraceptives and anti-inflammatory agents) and applications of physical methods (mass spectrometry, optical rotatory dispersion, magnetic circular dichroism) to organic and biochemical problems.

NAME	TITLE	BIRTHDATE (Mo., Day, Yr.)
LEDERBERG, JOSHUA	Professor and Executive Head, Department of Genetics	5-23-25
PLACE OF BIRTH (City, State, Country)	**PRESENT NATIONALITY** (If non-U.S. citizen, indicate kind of visa and expiration date)	**SEX**
Montclair, New Jersey	U.S.A.	☒ Male ☐ Female

EDUCATION: (Begin with baccalaureate training and include postdoctoral)

INSTITUTION AND LOCATION	DEGREE	YEAR CONFERRED	SCIENTIFIC FIELD
Columbia College, New York	B.A.	1944	
College of Physicians & Surgeons, Columbia University, New York (1944-46)			
Yale University	Ph.D.	1947	Microbiology

HONORS

1957 - National Academy of Sciences
1958 - Nobel Prize in Medicine

MAJOR RESEARCH INTEREST	ROLE IN PROPOSED PROJECT
Molecular Genetics; Artificial Intelligence	PRINCIPAL INVESTIGATOR

RESEARCH SUPPORT (See instructions)

SEE ATTACHMENTS:

RESEARCH AND/OR PROFESSIONAL EXPERIENCE (Starting with present position, list training and experience relevant to area of project. List all or most representative publications. Do not exceed 3 pages for each individual.)

1961-	Stanford University Director, Kennedy Laboratories for Molecular Medicine
1959-	Professor, Genetics and Biology, and Executive Head, Department of Genetics, Stanford University
1957-1959	University of Wisconsin Chairman, Department of Medical Genetics
1957	Melbourne University, Australia Fullbright Visiting Professor of Bacteriology
1950	University of California, Berkeley Visiting Professor of Bacteriology
1947-1959	University of Wisconsin Professor of Genetics
1946-1947	Yale University. Research Fellow of the Jane Coffin Childs Fund for Medical Research
1945-1946	Columbia University. Research Assistant in Zoology

Professional Activities:

1967-	NIMH: National Mental Health Advisory Council
1961-1962	President (Kennedy)'s Panel on Mental Retardation
1960-	NASA Committees: Lunar and Planetary Missions Board
1958-	National Academy of Sciences: Committees on Space Biology
1950-	President's Science Advisory Committee panels. National Institutes of Health, National Science Foundation study sections (genetics)

RESEARCH SUPPORT SUMMARY FOR JOSHUA LEDERBERG

	Grant Number	Grant Title	Current Year	Total Award	Grant Term	Budgeted % Time
1)	NASA:NGR-05-020-004	Cytochemical Studies of Planetary Micro-organisms	$150,000	$3,950,000	9/60-8/74 (Future support dubious)	4%
2)	NIH:AI-05160	Genetics of Bacteria	60,000	280,000	9/68-8/73 (Renewal Pending)	15%
3)	NIH:GM	Genetics Research Center	547,035	2,609,383	9/73-8/78 (Pending)	10%
4)	NIH:RR-00785	Stanford University Medical Experimental Computer Facility (SUMEX) Successor to #3	571,567	2,769,262	10/73-7/78	20%
5)	NIH: Computer Laboratory Health Care Resource Program	Large Scale Screening of Body Fluids for Metabolic Signs of Disease with Computer-managed Gas Chromatography and Mass Spectrometry	159,881	900,238	9/73-8/78 (Pending, Program funds impounded)	10%
6)	NIH:GM00295	Training Grant in Genetics	121,172	321,163	7/1/73-6/30/77	15%

SELECTED LIST OF PUBLICATIONS

Lederberg, J., 1959
A View of Genetics
Les Prix Nobel en 1958: 170-89.

Buchs, A., A. B. Delfino, A. M. Duffield, C. Djerassi, B. G. Buchanan, E. A. Feigenbaum, and J. Lederberg, 1970.
Applications of Artificial Intelligence for Chemical Inference. VI. Approach to a general method of interpreting low resolution mass spectra with a computer. Helvitia Chimica Acta 53 (6): 1394-1417.

Feigenbaum, E. A., B. G. Buchanan, J. Lederberg, 1971
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, P. 165-190.

Reynolds, W. E., V. A. Bacon, J. C. Bridges, T. C. Coburn, B. Halpern, J. Lederberg, E. C. Levinthal, E. Steed, R. B. Tucker, 1970
. A Computer Operated Mass Spectrometer System.
Analytical Chem. 42:1122-1129, September 1970.

Lederberg, J.
"Use of Computer to Identify Unknown Compounds: The Automation of Scientific Inference" in Biochemical Applications of Mass Spectrometry (G. R. Waller, ed.). John Wiley & Sons, New York (in press).

SECTION II – PRIVILEGED COMMUNICATION

BIOGRAPHICAL SKETCH

(Give the following information for all professional personnel listed on page 3, beginning with the Principal Investigator. Use continuation pages and follow the same general format for each person.)

NAME	TITLE	BIRTHDATE *(Mo., Day, Yr.)*
Feigenbaum, Edward A.	Principal Investigator, DENDRAL Project	1-20-36
PLACE OF BIRTH *(City, State, Country)*	**PRESENT NATIONALITY** *(If non-U.S. citizen, indicate kind of visa and expiration date)*	**SEX**
Weehawken, New Jersey	U.S. Citizen	☒ Male ☐ Female

EDUCATION *(Begin with baccalaureate training and include postdoctoral)*

INSTITUTION AND LOCATION	DEGREE	YEAR CONFERRED	SCIENTIFIC FIELD
Carnegie Institute of Technology Pittsburgh, Pennsylvania	B.S. Ph.D.	1956 1959	Electrical Engineering Behavioral Sciences.

HONORS and memberships:
American Psychological Association; Association for Computing Machinery (Member of the National Council 1966-68); American Association for the Advancement of Science, SIGBIO Chairman, 11/73-present.

MAJOR RESEARCH INTEREST	ROLE IN PROPOSED PROJECT
Artificial Intelligence	Principal Investigator

RESEARCH SUPPORT *(See instructions)*

RESEARCH AND/OR PROFESSIONAL EXPERIENCE *(Starting with present position, list training and experience relevant to area of project. List all or most representative publications. Do not exceed 3 pages for each individual.)*

1965- Stanford University, Computer Science Department Faculty
1965-1968 Stanford University, Director, Computation Center
1963 Summer Research Training Institute in Computer Simulation of Cognitive Processes (National Science Foundation)
1962 Carnegie Corporation. Summer Research Training Institute in Heuristic Programming. Faculty member.
1960-1964 University of California, Berkeley
Research-Center for Research in Management Science, 1960-1964
Research-Center for Human Learning, 1961-1964
Assistant and Associate Professor, School of Business Administration, 1960-64
1957-1960 The RAND Corporation, Santa Monica, California
1956 IBM Scientific Computing Center, New York

Selected Publications:

"Applications of Artificial Intelligence for Chemical Inference I. The Number of Possible Organic Compounds. Acyclic Structures Containing C, H, O and N", J. Am. Chem. Soc., 91, 2973 (1969). (Co-Author).

"Applications of Artificial Intelligence for Chemical Inference II. Interpretation of Low Resolution Mass Spectra of Ketones", J. Am. Chem. Soc., 91, 2977 (1969). (Co-Author).

Publications of Edward Feigenbaum

"Applications of Artificial Intelligence for Chemical Inference III. Aliphatic Ethers Diagnosed by their Low Resolution Mass Spectra and Nuclear Magnetic Resonance", J. Am. Chem. Soc., 91, 7440 (1969). (Co-Author).

"Heuristic DENDRAL: A Program for Generating Explanatory Hypotheses in Organic Chemistry", in Machine Intelligence 4, Edinburgh University Press, 1969. (Co-Author).

"Toward an Understanding of Information Processes of Scientific Inference in the Context of Organic Chemistry", in Machine Intelligence 5, Edinburgh University Press, 1970. (Co-Author).

"A Heuristic Program for Solving a Scientific Inference Problem: Summary of Motivation and Implementation", Stanford Artificial Intelligence Project Memo No. 104, November 1969. (Co-Author).

"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). (Co-Author).

"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). (Co-Author).

"Applications of Artificial Intelligence for Chemical Inference VI. Approach to a General Method of Interpreting Low Resolution Mass Spectra with a Computer", Chem. Acta Helvetica, 53, 1394 (1970). (Co-Author).

"On Generality and Problem Solving: A Case Study Using the DENDRAL Program", in Machine Intelligence 6, Edinburgh University Press (1971). (Co-Author).

"A Heuristic Programming Study of Theory Formation in Science", in proceedings of the Second International Joint Conference on Artificial Intelligence, Imperial College, London (September 1971). (Co-Author).

"Applications of Artificial Intelligence for Chemical Inference VIII. An Approach to the Computer Interpretation of the High Resolution Mass Spectra of Complex Molecules. Structure Elucidation of Estrogenic Steroids", Journal of the American Chemical Society, 94, 5962-5971 (1972). (Co-Author).

"Heuristic Theory Formation: Data Interpretation and Rule Formation", in Machine Intelligence 7, Edinburgh University Press (1972). (Co-Author).

"Applications of Artificial Intelligence for Chemical Inference X. Intsum A Data Interpretation Program as Applied to the Collected Mass Spectra of Estrogenic Steroids", Tetrahedron, 29, 3117 (1973). (co-author).

SECTION II – PRIVILEGED COMMUNICATION

BIOGRAPHICAL SKETCH

(Give the following information for all professional personnel listed on page 3, beginning with the Principal Investigator. Use continuation pages and follow the same general format for each person.)

NAME	TITLE	BIRTHDATE *(Mo., Day, Yr.)*
Buchanan, Bruce G.	Research Computer Scientist	7-7-40
PLACE OF BIRTH *(City, State, Country)*	**PRESENT NATIONALITY** *(If non-U.S. citizen, indicate kind of visa and expiration date)*	**SEX**
St. Louis, Missouri	U.S.Citizen	☒ Male ☐ Female

EDUCATION *(Begin with baccalaureate training and include postdoctoral)*

INSTITUTION AND LOCATION	DEGREE	YEAR CONFERRED	SCIENTIFIC FIELD
Ohio Wesleyan University	B.A.	1961	Mathematics
Michigan State University	M.A., Ph.D.	1966	Philosophy

HONORS

Recipient of National Institutes of Health Career Development Award (1971-1976)

MAJOR RESEARCH INTEREST	ROLE IN PROPOSED PROJECT
Artificial Intelligence	Associate Investigator

RESEARCH SUPPORT *(See instructions)*

NIH Research Career Development Award, GM-29662

RESEARCH AND/OR PROFESSIONAL EXPERIENCE *(Starting with present position, list training and experience relevant to area of project. List all or most representative publications. Do not exceed 3 pages for each individual.)*

1972-present Research Computer Scientist, Stanford University
1966-1971 Research Associate, Stanford Artificial Intelligence Project

Publications:

"On the Design of Inductive Systems: Some Philosophical Problems". British Journal for the Philosophy of Science 20 (1969), 311-323. (Co-Author).

"Applications of Artificial Intelligence for Chemical Inference II. Interpretation of Low Resolution Mass Spectra of Ketones". Journal of the American Chemical Society, 91, 2977-2981 (1969). (Co-Author).

"Applications of Artificial Intelligence for Chemical Inference I. The Number of Possible Organic Compounds: Acyclic Structures Containing C, H, O and N". Journal of the American Chemical Society, 91, 2973-2976 (1969). (Co-Author).

"Applications of Artificial Intelligence for Chemical Inference III. Aliphatic Ethers Diagnosed by Their Low Resolution Mass Spectra and NMR Data". Journal of the American Chemical Society, 91, 7440-45 (1969). (Co-Author).

"Heuristic DENDRAL: A Program for Generating Explanatory Hypotheses in Organic Chemistry". Machine Intelligence.4, Edinburgh University Press·(1969). (Co-Author).

Publications of Bruce Buchanan:

"Toward an Understanding of Information Processes of Scientific Inference in the Context of Organic Chemistry". Machine Intelligence 5, Edinburgh University Press (1969). (Co-Author).

"On Generality and Problem Solving: A Case Study Using the DENDRAL Program". Machine Intelligence 6, Edinburgh University Press (1969). (Co-Author).

"Some Speculation About Artificial Intelligence and Legal Reasoning". Stanford Law Review, Vol. 23, No. 1, November 1970. (Co-Author).

"Applications of Artificial Intelligence for Chemical Inference VI. Approach to a General Method of Interpreting Low Resolution Mass Spectra with a Computer". Chemica Acta Helvetica, 53, 1394 (1970). (Co-Author).

"An Application of Artificial Intelligence to the Interpretation of Mass Spectra". Mass Spectrometry Techniques and Appliances (1970).

"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, 93, 6831 (1970). (Co-Author).

"The Heuristic DENDRAL Program for Explaining Empirical Data". Proceedings of IFIP Congress 1971, Ljubljana, Yugoslavia. (Co-Author).

"A Heuristic Programming Study of Theory Formation in Science". Proceedings of Second International Joint Conference on Artificial Intelligence, Imperial College, London (1971). (Co-Author).

"Applications of Artificial Intelligence for Chemical Inference VIII. An Approach to the Computer Interpretation of the High Resolution Mass Spectra of Complex Molecules. Structure Elucidation of Estrogenic Steroids". Journal of the American Chemical Society, 1972. (Co-Author).

"Heuristic Theory Formation: Data Interpretation and Rule Formation". Machine Intelligence 7, Edinburgh University Press (1972). (Co-Author).

"Review of Hubert Dreyfus' 'What Computers Can't Do: A Critique of Artificial Reason'", Computing Reviews (January, 1973).

"Applications of Artificial Intelligence for Chemical Inference IX. Analysis of Mixtures Without Prior Separation as Illustrated for Estrogens". Submitted to the Journal of the American Chemical Society. (Co-Author).

"Applications of Artificial Intelligence for Chemical Inference X. Intsum A Data Interpretation Program as Applied to the Collected Mass Spectra of Estrogenic Steroids". Tetrahedron, 29, 3117 (1973). (co-author)

"Rule Formation on Non-Homogeneous Classes of Objects". In proceedings of the Third International Joint Conference on Artificial Intelligence (Stanford. 1973). (co-author).

"Current Status of the Heuristic DENDRAL Program for Applying Artificial Intelligence to the Interpretation of Mass Spectra". DENDRAL Project Memo, August 1973

Biographical Sketch of Bruce G. Buchanan

Memberships:

Association for Computing Machinery (ACM)
Philosophy of Science Association
American Association for Advancement of Science (AAAS)

Use continuation pages and follow the same general format for each person.)

NAME	TITLE	BIRTHDATE (Mo.-Day-Yr.)
Alan M. DUFFIELD	Research Associate	December 16 1936

PLACE OF BIRTH (City, State, Country)	PRESENT NATIONALITY (If non-U.S. citizen, indicate kind of visa and expiration date)	SEX
Perth, Western Australia	Australian, Permanent resident Immigrant Visa	☑ Male ☐ Female

EDUCATION (Begin with baccalaureate training and include postdoctoral)

INSTITUTION AND LOCATION	DEGREE	YEAR CONFERRED	SCIENTIFIC FIELD
University of Western Australia	B. Sc(1st Class Hons)	1958	Organic Chemistry
University of Western Australia	Ph.D.	1962	Organic Chemsitry

HONORS

MAJOR RESEARCH INTEREST	ROLE IN PROPOSED PROJECT
Applications of mass spectrometry to Biology and Biomedical Problems	Organic Chemist/mass spectroscopist

RESEARCH SUPPORT (See instructions)

N/A

RESEARCH AND/OR PROFESSIONAL EXPERIENCE (Starting with present position, list training and experience relevant to area of project. List all or most representative publications. Do not exceed 3 pages for each individual.)

1970 -	Research Associate, Department of Genetics, Stanford University School of Medicine
1969 -	Head of the Mass Spectrometry Laboratory, Chemistry Department Stanford University
1965 - 69	Research Associate, Department of Chemistry, Stanford University
1963 - 65	Postdoctoral Fellow, Department of Chemistry, Stanford University
1962 - 63	Postdoctoral Fellow, Department of Biochemistry, Stanford University School of Medicine.

PUBLICATIONS SINCE 1971

1. An Application of Artificial Intelligence to the Interpretation of Mass Spectra. Mass Spectrometry, B.W.G. Milne, Ed., John Wiley and Sons, New York, 1971, pp. 121-178
By B. G. Buchanan, A. M. Duffield and A. V. Robertson

2. Mass Spectrometry in Structural and Stereochemical Problems. CCIV. Spectra of Hydantoins.II. Electron Impact Induced Fragmentation of some Substituted Hydantoins.
   Org. Mass Spectr., 5, 551 (1971)
   By R. A. Corral, O. O. Orazi, A. M. Duffield and C. Djerassi

3. Electron Impact Induced Hydrogen Scrambling in Cyclohexanol and Isomeric Methylcyclohexanols.
   Org. Mass Spectr., 5, 383 (1971)
   By R. H. Shapiro, S. P. Levine and A. M. Duffield

4. Derivatives of 2-Biphenylcarboxylic Acid.
   Rev. Roumain. Chem., 16, 1095 (1971)
   By A. T. Balaban and A. M. Duffield

5. Alkaloide aus Evonymus europaea L.
   Helv. Chim. Acta, 54, 2144 (1971)
   By A. Klásek, T. Reichstein, A. M. Duffield and F. Santavý

6. Studies on Indian Medicinal Plants. XXVIII. Sesquiterpene Lactones of Enhydra Fluctuans Lour. Structures of Enhydrin, Fluctuanin and Fluctuadin.
   Tetrahedron, 28, 2285 (1972).
   By E. Ali, P. P. Ghosh Dastidar, S. C. Pakrashi, L. J. Durham and A. M. Duffield

7. The Electron Impact Promoted Fragmentation of Aurone Epoxides.
   Org. Mass Spectr., 6, 199 (1972)
   By B. A. Brady, W. I. O'Sullivan and A. M. Duffield

8. The Determination of Cyclohexylamine in Aqueous Solutions of Sodium Cyclamate by Electron Capture Gas Chromatography.
   Anal. Letters, 4, 301 (1971)
   By M. D. Soloman, W. E. Pereira and A. M. Duffield

9. Computer Recognition of Metastable Ions. Nineteenth Annual Conference on Mass Spectrometry, Atlanta, 1971, p. 63
   By A. M. Duffield, W. E. Reynolds, D. A. Anderson, R. A. Stillman, Jr. and C. E. Carroll

10. Spectrometrie de Masse. VI. Fragmentation de Dimethyl-2,2-dioxolanes-1,3-Insatures.
    Org. Mass Spectr., 5, 1409 (1971)
    By J. Kossanyi, J. Chuche and A. M. Duffield

11. Chlorpromazine Metabolism in Sheep. II. In vitro Metabolism and Preparation of 3H-7-Hydroxychlorpromazine.
    Journees D'Agressologie, 12, 333 (1971)
    By L. G. Brooks, M. A. Holmes, I. S. Forrest, V. A. Bacon, A. M. Duffield and M. D. Solomon

12. Mass Spectrometry in Structural and Stereochemical Problems. CCXVII. Electron Impact Promoted Fragmentation of O-Methyl Oximes of Some α,β-Unsaturated Ketones and Methyl Substituted Cyclohexanones.
    Canadian J. Chem., 50, 2776 (1972)
    By Y. M. Sheikh, R. J. Liedtke, A. M. Duffield and C. Djerassi

13. Thermal Fragmentation of Quinoline and Isoquinoline N-Oxides in the Ion Source of a Mass Spectrometer.
Acta Chem. Scand., 26, 2423 (1972).
By A. M. Duffield and O. Buchardt

14. Applications of Artificial Intelligence for Chemical Inference. VII. An Approach to the Computer Interpretation of the High Resolution Mass Spectra of Complex Molecules. Structure Elucidation of Estrogenic Steroids.
J. Amer. Chem. Soc., 94, 5962 (1972)
By D. H. Smith, B. G. Buchanan, R. S. Englemore, A. M. Duffield, A. Yeo, E. A. Feigenbaum, J. Lederberg and C. Djerassi

15. Mass Spectrometry in Structural and Stereochemical Problems. CCXIX. Identification of a Unidirectional Quadruple Hydrogen Transfer Process in 7-Phenyl-hept-3-en-2-one O-Methyl Oxime Ether.
Org. Mass Spectr., 6,1271 (1972).
By R. J. Liedtke, Y. M. Sheikh, A. M. Duffield and C. Djerassi

16. An Automated Gas Chromatographic Analysis of Phenylalanine in Serum.
Clinical Biochem., 5, 166 (1972)
By E. Steed, W. Pereira, B. Halpern, M. D. Solomon and A. M. Duffield

17. Pyrrolizidine Alkaloids. XIX. Structure of the Alkaloid Erucifoline.
Coll. Czech. Chem. Commun., (1972)
By P. Sedmera, A. Klásek, A. M. Duffield and F. Santavý.

18. Mass Spectrometry in Structural and Stereochemical Problems. CCXXII. Delineation of Competing Fragmentation Pathways of Complex Molecules from a Study of Metastable Ion Transitions of Deuterated Derivatives.
Org. Mass Spectr., 7, (1973)
By D. H. Smith, A. M. Duffield and C. Djerassi

19. Chlorination Studies I. The Reaction of Aqueous Hypochlorous Acid with Cytosine.
Biochem. Biophys. Res. Commun., 48, 880 (1972)
By W. Patton, V. Bacon, A. M. Duffield, B. Halpern, Y. Hoyano, W. Pereira and J. Lederberg

20. A Study of the Electron Impact Fragmentation of Promazine Sulphoxide and Promazine using Specifically Deuterated Analogues.
Austral. J. Chem., 26, (1973).
By M. D. Solomon, R. Summons, W. Pereira and A. M. Duffield

21. Spectrometric de Masse. VIII. Elimination d'eau Induite par Impact Electronique dans le Tetrhydro-1,2,3,4-naphtalenediol-1,2.
Org. Mass. Spectrom., 7 (1973).
By P. Perros, J. P. Morizui, J. Kossanyi and A. M. Duffield

22. The Determination of Phenylalanine in Serum by Mass Fragmentography
Clinical Biochem., submitted for publication (1973).
By W. E. Pereira, V. A. Bacon, Y. Hoyano, R. Summons and A. M. Duffield

SECTION II – PRIVILEGED COMMUNICATION

BIOGRAPHICAL SKETCH

(Give the following information for all professional personnel listed on page 3, beginning with the Principal Investigator. Use continuation pages and follow the same general format for each person.)

NAME	TITLE	BIRTHDATE *(Mo., Day, Yr.)*
Dennis H. Smith	Research Associate	11/12/42

PLACE OF BIRTH *(City, State, Country)*	PRESENT NATIONALITY *(If non-U.S. citizen, indicate kind of visa and expiration date)*	SEX
New York	USA	X Male ☐ Female

EDUCATION *(Begin with baccalaureate training and include postdoctoral)*

INSTITUTION AND LOCATION	DEGREE	YEAR CONFERRED	SCIENTIFIC FIELD
Massachusetts Inst. of Technology Cambridge, Mass.	S.B.	1964	Chemistry
University of California, Berkeley Berkeley, California	Ph.D.	1967	Chemistry

HONORS

Alfred P. Sloan Foundation Scholarship
NASA Predoctoral Traineeship
Phi Lambda Upsilon, Sigma Xi

MAJOR RESEARCH INTEREST	ROLE IN PROPOSED PROJECT
Mass Spectrometry and A.I. in Chemistry	Research Associate

RESEARCH SUPPORT *(See instructions)*

N/A

RESEARCH AND/OR PROFESSIONAL EXPERIENCE *(Starting with present position, list training and experience relevant to area of project. List all or most representative publications. Do not exceed 3 pages for each individual.)*

1971-Present Research Associate, Stanford University, Stanford,Ca.
1970-1971 Visiting Scientist, University of Bristol, Bristol, England
1967-1970 Assistant Research Chemist, University of Calif.at Berkeley, Berkeley, Ca.
1965-1967 NASA Pre-Doctoral Traineeship, University of Calif.at Berkeley,Berkeley, Ca.

Publications: See attached list.

PUBLICATIONS: D. H. SMITH

1. H. G. Langer, R. S. Gohlke and D. H. Smith, "Mass Spectrometric Differential Thermal Analysis," Anal. Chem., 37, 433 (1965).

2. S. M. Kupchan, J. M. Cassady, J. E. Kelsey, H. K. Schnoes, D. H. Smith and A. L. Burlingame, "Structural Elucidation and High Resolution Mass Spectrometry of Gaillardin, a New Cytotoxic Sesquiterpene Lactone," J. Amer. Chem. Soc., 88, 5292 (1966).

3. D. H. Smith, Ph.D. Thesis, "High Resolution Mass Spectrometry: Techniques and Applications to Molecular Structure Problems," Dept. of Chemistry, University of California, Berkeley, California (1967).

4. H. K. Schnoes, D. H. Smith, A. L. Burlingame, P. W. Jeffs and W. Döpke, "Mass Spectra of Amaryllidaceae Alkaloids: The Lycorenine Series," Tetrahedron, 24, 2825 (1968).

5. A. L. Burlingame, D. H. Smith and R. W. Olsen, "High Resolution Mass Spectrometry in Molecular Structure Studies, XIV. Real-time Data Acquisition, Processing and Display of High Resolution Mass Spectral Data," Anal. Chem., 40, 13 (1968).

6. A. L. Burlingame and D. H. Smith, "High Resolution Mass Spectrometry in Molecular Structure Studies II. Automated Heteroatomic Plotting as an Aid to the Presentation and Interpretation of High Resolution Mass Spectral Data," Tetrahedron, 24, 5749 (1968).

7. W. J. Richter, B. R. Simoneit, D. H. Smith and A. L. Burlingame, "Detection and Identification of Oxocarboxylic and Dicarboxylic Acids in Complex Mixtures by Reductive Silylation and Computer-Aided Analysis of High Resolution Mass Spectral Data," Anal. Chem., 41, 1392 (1969).

8. The Lunar Sample Preliminary Examination Team, "Preliminary Examination of Lunar Samples from Apollo 11," Science, 165, 1211 (1969).

9. S. M. Kupchan, W. K. Anderson, P. Bollinger, R. W. Doskotch, R. M. Smith, J. A. Saenz Renauld, H. K. Schnoes, A. L. Burlingame and D. H. Smith, "Tumor Inhibitors, XXXIX. Active Principles of Acnistus arborescens. Isolation and Structural and Spectral Studies of Withaferin A and Withacnistin," J. Org. Chem., 34, 3858 (1969).

10. A. L. Burlingame, D. H. Smith, T. O. Merren and R. W. Olsen, "Real-time High Resolution Mass Spectrometry," in Computers in Analytical Chemistry (Vol. 4 in Progress in Analytical Chemistry series), C. H. Orr and J. Norris, Eds., Plenum Press, New York, 1970, pp. 17-38.

11. The Lunar Sample Preliminary Examination Team, "Preliminary Examination of Lunar Samples from Apollo 12," Science, 167, 1325 (1970).

12. D. H. Smith, "Mass Spectrometry," Chapter X in Guide to Modern Methods of Instrumental Analysis, T. M. Gouw, Ed., Wiley-Interscience, New York, 1972.

13. D. H. Smith, R. W. Olsen, F. C. Walls and A. L. Burlingame, "Real-time Mass Spectrometry: LOGOS--A Generalized Mass Spectrometry Computer System for High and Low Resolution, GC/MS and Closed-Loop Applications," Anal. Chem., 43, 1796 (1971).

14. A. L. Burlingame, J. S. Hauser, B. R. Simoneit, D. H. Smith, K. Biemann, N. Mancuso, R. Murphy, D. A. Flory and M. A. Reynolds, "Preliminary Organic Analysis of the Apollo 12 Cores," Proceedings of the Apollo 12 Lunar Science Conference, E. Levinson, Ed., M.I.T. Press, Cambridge, Mass., 1971, p. 1891.

15. D. H. Smith, "A Compound Classifier Based on Computer Analysis of Low Resolution Mass Spectral Data," Anal. Chem., 44, 536 (1972).

16. D. H. Smith and G. Eglinton, "Compound Classification by Computer Treatment of Low Resolution Mass Spectra-Application to Geochemical and Environmental Problems," Nature, 235, 325 (1972).

17. D. H. Smith, N. A. B. Gray, C. T. Pillinger, B. J. Kimble and G. Eglinton, "Complex Mixture Analysis - Geochemical and Environmental Applications of a Compound Classifier Based on Computer Analysis of Low Resolution Mass Spectra," Adv. in Org. Geochem., 1971, p. 249.

18. D. H. Smith, B. G. Buchanan, R. S. Engelmore, A. M. Duffield, A. Yeo, E. A. Feigenbaum, J. Lederberg and C. Djerassi, "Applications of Artificial Intelligence for Chemical Inference, VIII. An Approach to the Computer Interpretation of the High Resolution Mass Spectra of Complex Molecules. Structure Elucidation of Estrogenic Steroids," J. Amer. Chem. Soc., 94, 5962 (1972).

19. D. H. Smith, A. M. Duffield and C. Djerassi, "Mass Spectrometry in Structural and Stereochemical Problems, CCXXII. Delineation of Competing Fragmentation Pathways of Complex Molecules from a Study of Metastable Ion Transitions of Deuterated Derivatives," Org. Mass. Spectrom., 7, 367 (1973).

20. P. Longevialle, D. H. Smith, H. M. Fales, R. J. Highet and A. L. Burlingame, "High Resolution Mass Spectrometry in Molecular Structure Studies, V. The Fragmentation of Amaryllis Alkaloids in the Crinine Series," Org. Mass Spectrom., 7, 401 (1973).

21. B. R. Simoneit, D. H. Smith, G. Eglinton and A. L. Burlingame. "Applications of Real-time Mass Spectrometric Techniques to Environmental Organic Geochemistry, II. San Francisco Bay Area Waters," Arch. Env. Contam and Tox., 1, 193 (1973).

22. D. H. Smith, B. G. Buchanan, R. S. Engelmore, H. Adlercreutz and C. Djerassi, "Applications of Artificial Intelligence for Chemical inference, IX. Analysis of Mixtures Without Prior Separation as Illustrated for Estrogens," J. Amer. Chem. Soc., 95, 6078 (1973).

23. D. H. Smith, B. G. Buchanan, W. C. White, E. A. Feigenbaum, J. Lederberg and C. Djerassi, "Applications of Artificial Intelligence for Chemical Inference, X. INTSUM - A Data Interpretation and Summary Program Applied to the Collected Mass Spectra of Estrogenic Steroids," Tetrahedron, 29, 3117 (1973).

24. G. Loew, M. Chadwick and D. H. Smith, "Applications of Molecular Orbital Theory to the Interpretation of Mass Spectra. Prediction of Primary Fragmentation Sites in Organic Molecules," Org. Mass Spectrom., 7, 1241 (1973).

SECTION II – PRIVILEGED COMMUNICATION

BIOGRAPHICAL SKETCH

(Give the following information for all professional personnel listed on page 3, beginning with the Principal Investigator. Use continuation pages and follow the same general format for each person.)

NAME	TITLE	BIRTHDATE (Mo., Day, Yr.)
Sridharan, Natesa S.	Research Associate	10/2/46
PLACE OF BIRTH (City, State, Country)	**PRESENT NATIONALITY (If non-U.S. citizen, indicate kind of visa and expiration date)**	**SEX**
Madras, India	India; 5/73-U.S. permanent residence	☒ Male ☐ Female

EDUCATION *(Begin with baccalaureate training and include postdoctoral)*

INSTITUTION AND LOCATION	DEGREE	YEAR CONFERRED	SCIENTIFIC FIELD
Indian Institute of Technology, Madras India	Bachelor of Technology	1967	Electrical Engineering
State University of New York, Stony Brook	M.S.	1969	Computer Science
	Ph.D.	1971	Computer Science

HONORS

University Fellow - 1968-1971, SUNY Stony Brook; Graduate Assistant - 1967-1968, SUNY Stony Brook; Siemens' Award (awarded for top rank in Electrical Engineering) - 1967, ITT Madras; National Merit Scholarship - 1963-1967, ITT Madras

MAJOR RESEARCH INTEREST	ROLE IN PROPOSED PROJECT
Computer Application in Chemistry and Medicine	Research Associate

RESEARCH SUPPORT *(See instructions)*

RESEARCH AND/OR PROFESSIONAL EXPERIENCE *(Starting with present position, list training and experience relevant to area of project. List all or most representative publications. Do not exceed 3 pages for each individual.)*

1971-present Research Associate, Heuristic Programming Project, Stanford University
1970-1971 Consultant, IAC Computer Corp., Long Island, N.Y.

Sridharan, N.S., "An Application of Artificial Intelligence to Organic Chemical Synthesis" Doctoral Thesis, State University of New York at StonyBrook, 1971.

Sridharan, N.S., "Search Strategies of Organic Chemical Synthesis", Third International Joint Conference on Artificial Intelligence (3IJCAI), Stanford, 1973

Sridharan, N.S. (co-author), "Heuristic DENDRAL: Analysis of Molecular Structure", Proc. NATO Advanced Study Institute, Amsterdam, 1973.

Sridharan, N.S. (co-author), "Heuristic Theory Formation", Machine Intelligence, Volume 7, Edinburgh, 1972.

SECTION II – PRIVILEGED COMMUNICATION

BIOGRAPHICAL SKETCH

(Give the following information for all professional personnel listed on page 3, beginning with the Principal Investigator. Use continuation pages and follow the same general format for each person.)

NAME	TITLE	BIRTHDATE *(Mo., Day, Yr.)*
Brown, Harold D.	Associate Professor	July 12,1934
PLACE OF BIRTH *(City, State, Country)*	**PRESENT NATIONALITY** *(If non-U.S. citizen, indicate kind of visa and expiration date)*	**SEX**
South Bend, Indiana	U.S.	☒ Male ☐ Female

EDUCATION *(Begin with baccalaureate training and include postdoctoral)*

INSTITUTION AND LOCATION	DEGREE	YEAR CONFERRED	SCIENTIFIC FIELD
University of Notre Dame, Notre Dame, Indiana	M.Sc.	1963	Mathematics
Ohio State University, Columbus, Ohio	Ph.D.	1966	Mathamatics
(No Baccalaureate Degree)			

HONORS

Summa Cum Laude - Notre Dame

MAJOR RESEARCH INTEREST	ROLE IN PROPOSED PROJECT
Applied Discrete Mathematics - Computer Science	Research Associate

RESEARCH SUPPORT *(See instructions)*

Principal Investigator, NSF-GP-16793 (Expires March, 1974)

Pending Proposal NSF (Proposed starting date September, 1974)

RESEARCH AND/OR PROFESSIONAL EXPERIENCE *(Starting with present position, list training and experience relevant to area of project. List all or most representative publications. Do not exceed 3 pages for each individual.)*

Visiting Associate Professor, Computer Science, Stanford University , 1971-72, 1973-present
Associate Professor, Mathematics, Ohio State University, 1966-
Visiting Professor, Mathematics, Rhine,Westf. Tech. Hoch., Aachen, 1972 and 1973
Visiting Member, Courant Institute, New York University, 1967-68
Instructor/Assistant Professor, Assistant Chairman, Mathematics, Ohio State U., 1963-66
Assistant to the Chairman, Mathematics, University of Notre Dame, 1960-63
Director or Associate Director, NSF-SSTP, 1964-70

Publications

Near Algebras, Ill. J. Math. 12(1968), Pg. 215.

Distributor Theory in Near Algebras, Comm. Pure App. Math. XXI(1968), Pg. 535.

An Algorithm for the Determination of Space Groups, Math. Comp. 23(1969), Pg. 499.

Some Empirical Observations on Primitive Roots, with H. Zassenhaus, J. Number Theory 3(1971), Pg. 306.

A Generalization of Farey Sequences, with K. Mahler, J. Number Theory 3(1971), pg.364.

Basic Computations for Orders, Stanford CS Memo STAN-CS-72-208.

An Application of Zassenhaus' Unit Theorem, Acta Arith. XX(1972), Pg. 154.

Integral Groups I: The Reducible case, with J. Neubüser and H. Zassenhaus, Numer. Math. 19(1972), Pg. 386.

Integral Groups II: The Irreducible Case, with J. Neubüser and H. Zassenhaus, Numer. Math. 20(1972), Pg. 22.

Integral Groups III: Normalizers, with J. Neubüser and H. Zassenhaus, Math. Comp. 27(1973), Pg. 167.

Constructive Graph Labeling Via Double Cosets, with L. Hjelmeland and L. Masinter, Discrete Math. in press and Stanford CS Memo STAN-CS-72-318.

An Algorithm for the Construction of the Graphs of Organic Molecules, with L. Masinter, Discrete Math. in press and Stanford CS Memo STAN-CS-73-261.

The Crystallographic Groups of 4-dimensional Space, with J. Neubüser, H. Wondratschek and H. Zassenhaus, Wiley-Interscience in press.

SECTION II – PRIVILEGED COMMUNICATION

BIOGRAPHICAL SKETCH

(Give the following information for all professional personnel listed on page 3, beginning with the Principal Investigator. Use continuation pages and follow the same general format for each person.)

NAME	TITLE	BIRTHDATE (Mo., Day, Yr.)
DROMEY, Robert Geoffrey	Research Associate	11/21/46
PLACE OF BIRTH (City, State, Country)	**PRESENT NATIONALITY** (If non-U.S. citizen, indicate kind of visa and expiration date)	**SEX**
Castlemaine, Victoria, Australia	Australian, J-1 Visa, Exp. 10/8/74	☒ Male ☐ Female

EDUCATION *(Begin with baccalaureate training and include postdoctoral)*

INSTITUTION AND LOCATION	DEGREE	YEAR CONFERRED	SCIENTIFIC FIELD
Swinburne College of Technology, Melbourne, Australia	Diploma of Appl. Chem.	1968	Chemistry
La Trobe University Melbourne, Australia	Ph.D.	1973	Molecular Science

HONORS CSIRO Postdoctoral Studentship
Commonwealth Postgraduate Research Scholarship
Walter Lindrum Memorial Scholarship
Equivalent of First Class Honors Master of Science Preliminary (1969)

MAJOR RESEARCH INTEREST	ROLE IN PROPOSED PROJECT
Application of Artificial Intelligence Techniques to Bio-Medical and Chemical Problems.	Research Associate

RESEARCH SUPPORT *(See instructions)*

RESEARCH AND/OR PROFESSIONAL EXPERIENCE *(Starting with present position, list training and experience relevant to area of project. List all or most representative publications. Do not exceed 3 pages for each individual.)*

1973 DENDRAL Project, Stanford University, Computer Science Department
1973 Software Development for Graphics Systems, LaTrobe University, Computer Centre
1969-73 Construction, development and applications of an on-line photoelectron spectrometer LaTrobe University, Chemistry Department
1969-73 Application of Deconvolution Techniques to the Processing of Experimental Data.

Publications:

"Deconvolution and Its Application to the Processing of Experimental Data", Intl. Journal of Mass Spectrometry and Ion Physics, 1970, 4. (co-author).

"Inverse Convolution in Mass Spectrometry", Intl. Jnl. Mass Spec. Ion Phys.,1971, 6. (co-author).

"A Combined Time Averaging-Deconvolution Technique Applied to Electron Impact Ionization Efficiency Curves", Internation Journal of Mass Spectrometry & Ion Physics, 1971, 6. (co-author).

"The Perfect Direction and Velocity Focus at 254°34' in a Cylindrical Electrostatic Field", Reviews of Scientific Instruments, 1973, 44. (co-author).

"Detection of Spin-Orbit Splitting in the Photoelectron Spectrum of O_2^+ by Deconvolution", Chem. Physics Letters (in press), 1973. (co-author)

"The Effect of Finite Line Widths on the Interpretation of Photoelectron Spectra", Journal of Electron Spectroscopic (accepted for publication). (co-author).

"An On-line Ultraviolet Photoelectron Spectrometer for High-Resolution Studies of Molecular Structure", Australian Journal of Chemistry (accepted for publication). (co-author).

"Photoelectron Spectroscopic Correlation of the Molecular Orbitals of the Alkanes and Alkyliodides", Journal of Molecular Structure (submitted for publication). (coauthor).

"Comparison of the Photoelectron Spectra and the Photoionization Efficiency Curves for the Alkyliodides", Transactions of the Faraday Society (submitted for publication). (co-author).

"A Convolution-Deconvolution Algorithm Using Fast Fourier Transforms), Decuscope, 1973 (in press).

RESEARCH PLAN

BIOMOLECULAR CHARACTERIZATION: ARTIFICIAL INTELLIGENCE
A Program of Resource-Related Research

I. INTRODUCTION

This renewal application is intended to sustain and augment the capabilities of the mass spectrometry (MS) program which has served as a major institutional resource at Stanford for some years. With previous support from NASA and NSF it has made possible a highly interdisciplinary set of research projects ranging over: artificial intelligence (AI) in biomolecular characterization, natural product chemistry, clinical biochemical studies on steroids, and the mechanisms of molecular fragment formation in mass spectrometry. While the facility equipment for mass spectrometry has been funded mostly by other agencies, connected research programs embrace several NIH research projects as well. In addition, this activity was closely coupled with the ACME Medical School computer resource (1966-1973) and will have similar associations with the new AIM-SUMEX computer resource recently funded by the BRB (see Section I.C).

Previous support reflects the diversified facets of this interdisciplinary research. NASA has supported projects in new instrumentation, including the initial mass spectrometer-computer link, NSF has supported chemical research, and ARPA has supported our artificial intelligence research and initial application to mass spectrometry. Overall cutbacks have forced NASA to reduce funding for this area of research despite their interest. Under ARPA support to Drs. Feigenbaum and Lederberg for AI research, the DENDRAL program became recognized as one of the most successful AI applications programs. However, ARPA is chartered to fund frontier computer science research and no longer provides funds for the DENDRAL applications programs. ARPA has indicated a reluctance to continue funding to this group for the theory formation work in chemistry, although we expect to continue to receive ARPA support for more theoretical aspects of our research program (e.g., automatic programming).

We previously submitted a comprehensive proposal to the NIH (RR-00785, 3/28/73) which included an application for the AIM-SUMEX computing resource and a renewal of the existing DENDRAL grant (RR-00612). This proposal was approved for 5 years by the National Advisory Research Resources Council. Certain reservations were, however, communicated to us: they concerned especially what we must agree was an ambitious effort to close the control loop for "intelligent automation" whose costs overreached the immediate utility of the expected result. During subsequent discussions with the Biotechnology Resources Branch, taking into account the council review and a number of diverse policy issues, we agreed administratively to segment the two components of the original proposal. The AIM-SUMEX portion of the original proposal (excluding DENDRAL) was recently funded for 5 years as a national resource for artificial intelligence in medicine. The present proposal for resource-related research in biomolecular characterization and artificial intelligence is an elaboration of the DENDRAL portion incorporating intensive reexamination and revision of the previous proposal.

With the differentiation of priorities represented by AIM-SUMEX,

the Genetics Research Center (GRC), and continuing work on artificial intelligence under Dr. Feigenbaum's leadership, the present renewal application places more emphasis than heretofore on real-world oriented applications. Correspondingly, we have agreed that it is now more appropriate that Dr. Djerassi should be designated as Principal Investigator in this phase of our work.

As outlined in section B.2, the interests and responsibilities of Professors Djerassi (Chemistry), Feigenbaum (Computer Science) and Lederberg (Genetics) have been closely interdigitated. With their further connections with many colleagues, these programs enjoy a high degree of university-wide participation. For example, the Genetics Department is also closely affiliated with Biology, Biochemistry, Pediatrics, Psychiatry and Medicine through joint appointments or joint research projects or both. This breadth would be difficult to obtain except at a few institutions where the medical school is both academically and geographically integrated with the university to the degree that characterizes the Stanford University environment.

GLOSSARY OF ABBREVIATIONS

ACME	- Advanced Computer for Medical Research (Nih-funded computer resource, 1968-1973)
AI	- artificial intelligence
AIM-SUMEX	- A comprehensive computer resource intended to serve the national requirement for artificial intelligence in medicine. This will be implemented at the Stanford University facility called AIM-SUMEX
ARPA	- Advanced Research Projects Agency of the Department of Defense.
BRB	- Biotechnology Resources Branch
13CMR	- carbon-13 magnetic resonance
GC	- gas chromatography or gas chromatograph
GRC	- Genetics Research Center (Stanford, J. Lederberg, Principal Investigator; NIGMS-approved and awaiting funding. Grant #P01-GM 20832-01)
HRMS	- high resolution mass spectrometry
IR	- infra-red
IRL	- Instrumentation Research Laboratory (Stanford Genetics Department)
LRMS	- low resolution mass spectrometry
MCD	- magnetic circular dichroism
MS	- Mass spectrometry or mass spectrometer
NASA	- National Aeronautics & Space Administration
NMR	- nuclear magnetic resonance
NSF	- National Science Foundation
ORD	- optical rotatory dispersion
PL/ACME	- a modified version of the PL-1 computer language (for the Stanford ACME computer facility)
SUMEX	- Stanford University Medical Experimental Computer Resource (NIH funded computer resource, 1973-1978)
UV	- ultra-violet

A. OBJECTIVES:

Core Research.
The funds now applied for would permit

1) the continued funding of the MS laboratory as a biomolecular characterization resource;

2) advancement of laboratory instrumentation capability in specific areas of GC-HRMS and the exploitation of metastable peak analysis.

3) the further development of AI computer techniques to match the instrumentation. This work will emphasize practical utilization for applications in biomolecular characterization connected with other on-going biomedical research programs. It will include, for example, a) the analysis of mixtures by GC/MS; b) metastable peak analysis for difficult problems of pure compounds and of mixtures not readily separable by GC; c) optimized data analysis for characterization of MS peaks and d) heuristic analysis of spectra for the molecular ion composition.

Our project is the only systematic effort, to our knowledge, currently underway in this country for computer assisted structure elucidation. Subsequent to our early publications, an intensive program has been mounted in Japan in similar areas. This situation may be contrasted with computer assisted organic synthesis, an area receiving considerable attention from several research groups. These capabilities can be beneficially provided to a wider community via the AIM-SUMEX resource. Research on the emulation of human intellect by computer programs will undoubtedly influence the efficiency with which chemical research can be applied to ever more complex problems of health, e.g., intermediary metabolism and its pathologies; environmental influences on health; the development and critical validation of new therapeutic agents.

The achievement of these objectives depends on the continued maintenance and development of the DENDRAL AI programming system (see below). The advent of the AIM-SUMEX facility will remove some of the serious computational limits on the exercise of this system that have delayed recent progress.

Education.
In our university setting, pre-doctoral and post-doctoral education of course constitutes a part of our mission. As far as is practically possible, research participation in the DENDRAL program has been coupled with dissertation work by graduate students and post-doctoral research experience respectively. Examples of people (and their research area) whose education has been enhanced in this way are the following:

Graduate Students: J. Simek, pedagogical aspects of the structure generator; Wai Lee Tan, synthesis of new estrogen compounds; H. Eggert, 13CMR of amines and steroidal ketones; C. Van Antwerp, 13CMR of steroidal alcohols; C. Farrell, theory formation from mass spectral data; L. Masinter, development of the structure generator; M. Stefik, AI applications to chemistry.

Postdoctoral Fellows: G. Dromey, theory formation from analytical data; R. Gritter, mass spectral fragmentation of biologically active steroids; R. Carhart, analysis of 13CMR spectra by DENDRAL-like programs; S. Hammerum, development of better fragmentation rules for progesterones.

Formal organization.
This project has been a long-term commitment of Djerassi, Lederberg and Feigenbaum functioning in effect as co-investigators. We coordinate our activities with day-to-day contacts in the pursuit of convergent research objectives. In the light of the extension of our collaborative activity during the last two years, we are now organizing a formal advisory group to include, in addition to ourselves, H. Cann, J. Barchas, and E. Van Tamelen. This group will advise the principal investigators on the direction of the program with respect to allocating available facilities and seeking out and helping other collaborators. This designation simply recognizes the fact that many of our colleagues have already been engaged in relevant collaborative research with us. A MS resource has recently been funded at the University of California/Berkeley, under the direction of Dr. A.L. Burlingame. Drs. Djerassi and Burlingame have recently engaged in some collaborative research which was made more successful by the sharing of facilities and expertise available at one institution but not at the other. We would hope to maintain and strengthen these contacts to avoid unnecessary duplication of effort.

We plan to discuss with Dr. A.L. Burlingame the most appropriate procedures for coordinating the related activities of our respective programs at the University of California/Berkeley and here. This may take the form of reciprocal membership in advisory committees.

The "hardware resource" to which this application is pegged has been identified as the MS facility. While these instruments alone represent an investment of over $300,000, funded previously by several agencies, they do not represent the most important resource. We would use this designation instead for the working team led by the principal and co-investigators. The skills embraced by this group include, as mentioned, computer science, structural organic chemistry, molecular biology, instrumentation engineering and a wide range of other disciplines. They are represented not only in the principal professors but in a diversified and accomplished professional research staff (see Budget Justification). The program for which funds are now requested is the vital means by which the interests of this group can be sustained in a coordinated effort that would be very costly both in funds and in time if it had to be reconstructed from scratch. Without the financial support now requested, this line of collaborative research will have to be abandoned, with it a unique style of interdisciplinary collaboration, and the MS facility will be terminated.

B. BACKGROUND AND RATIONALE

1. The Structure Elucidation Problem

a) The General Problem. Analysis of molecular structure is a major activity in our program of resource related research. For the specific task of elucidating molecular structures, i.e., the topology of atom-to-atom connectivities, analysts utilize a mixture of information derived from chemical procedures and spectroscopic techniques. Each item of information, if not redundant or uninterpretable, contributes to the solution of the problem. Chemists draw upon a tremendous body of specific knowledge about the task area (e.g., clinical chemistry, biochemistry), molecular structure, spectroscopic techniques, etc., in order to piece together this information and infer the structure of molecules. These features, and the relative simplicity of the final concept of a structure, make the problem particularly well-suited for applications of the techniques of AI to assist research workers performing the task.

b) Djerassi's Laboratory. Professor Djerassi has been concerned with structure elucidation problems since the beginning of his chemical research. His activities at Stanford have been concerned heavily with the application of particular spectroscopic techniques to structural studies of biomedically important compounds. These techniques include optical rotatory dispersion (ORD) and, more recently, magnetic circular dichroism (MCD) (both of them supported initially by the NIH). Since 1961 he and his group have also been concerned with MS because of the power of the technique, in terms of specificity and sensitivity, as an analytical tool for structure elucidation. Four books and approximately 250 articles on MS have been published by him and his colleagues.

The technique of MS does not suffice for all structure determination problems, but it is a very powerful tool in areas where there exists a body of knowledge about the MS behavior of related molecules. When sample size is limited MS may well be the only technique that can be utilized. The recent availability of high resolution mass spectrometers has made HRMS the technique of choice for many applications because under ideal conditions the exact mass number uniquely specifies the the empirical formula of a molecule or fragment. On a parallel course, the technique of GC/MS, routinely available with low resolution mass spectrometers (GC/LRMS), has revolutionized investigations wherever complex mixtures are encountered. All of the above considerations argue that an extension of MS at Stanford to provide routine GC/LRMS and GC/HRMS analyses would be the next logical step to assist researchers depending on this facility for solutions of their structure elucidation problems.

2. Historical Background

a) Mass Spectrometry Laboratory. Prior to the existing DENDRAL grant, the groundwork was laid for computerization of the existing mass spectrometers, an Associated Electrical Industries MS-9 high resolution mass spectrometer and an Atlas CH-4 low resolution mass spectrometer. This work, supported primarily by NASA via the

Instrumentation Research Laboratory (IRL) in the Department of Genetics, resulted in link-up to the then existing ACME computer facility via a PDP-11 mini-computer which acted as a buffer between the spectrometers and ACME. Initial data acquisition and reduction programs were written for the system and utilized on a limited basis. The funding of the DENDRAL proposal, NIH grant RR-612 (May 1,1971-present) in conjunction with additional resources provided by the IRL resulted in a major improvement to these capabilities. The fruits of these efforts are described under section I.B.3 (below).

b) Summary of Early DENDRAL Development.
In 1964, Lederberg devised a notational algorithm for chemical structures (termed DENDRAL) that allowed questions of molecular structure to be framed in precise graph-theoretic terms. (Refs. 1,3-5,12). He also showed how to use the DENDRAL algorithm to generate complete and irredundant lists of structural isomers. (Refs. 1,6).

In 1965-66 Lederberg and Feigenbaum began exploring the idea of using the isomer generator in an artificial intelligence program - searching the space of possible structures for plausible solutions to a problem much as a chess-playing program searches the space of legal moves for the best moves. (Refs. 7,12). This approach guarantees that every possible solution to a problem is considered - either implicitly, as when whole classes of unstable structures are rejected, or explicitly, as when complete molecules are tested for plausibility. In either case, an investigator easily determines the criteria for rejection and acceptance and knows that no possibilities have been forgotten. This approach also guarantees that structures appear in the list only once - that automorphic representations of the same complex molecule have not been included. In both these respects the computer program has an advantage over manual approaches to structure elucidation.

c) Initial collaboration with Djerassi. (Refs. 14,15,19, 20,21,22,24).
Lederberg and Feigenbaum realized that (a) only through application to real problems could the AI approach be materially advanced and critically evaluated, and (b) MS appeared to be a fruitful applications area. MS appeared to be an excellent problem area because of the close relationship between spectral fragmentation patterns and molecular structure for many classes of molecules. Djerassi's interest and expertise - and daily interaction between members of his group and the AI group - led to a series of joint publications describing the approach and initial results of the programs. The success of these collaborative efforts led to the proposal to the NIH for initial funding to extend these efforts.

d) Efforts Under NIH Funding for DENDRAL. (Refs. 25-41).
The initial funding by NIH provided the opportunity to upgrade the instrumentation and computer programs. In particular we were able to mount a concerted project on both the analysis of mass spectra of biomedically important compounds and the mathematical aspects of molecular structure. Progress reports to the NIH describe this research in detail. The most recent annual report appears in Appendix B. A series of publications directed to audiences both in computer science and chemistry are listed in the bibliography. The following section (Section 3) summarizes the capabilities for

structure elucidation which, in themselves, constitute an important result of past work.

e) Related Research.

An important side effect of the DENDRAL project is the extent to which additional research was inspired and carried out to fill gaps in existing knowledge. This research, not supported by the DENDRAL grant, has been beneficial to on-going DENDRAL work, and vice-versa. Publications which have arisen from this research are listed in the bibliography (Refs. 58-70). A brief review of these publications should indicate the need for precise specification of the knowledge elicited from chemists and used in computer programs. As an example, consider the description and application of an early algorithm for generation of cyclic structural isomers (21). This paper considered the problem of spectroscopic differentiation of isomers of C6H10O. Unsaturated ethers fall in one of the classes of isomeric compounds which must be considered, but the MS of unsaturated ethers had not been investigated systematically. This work was subsequently carried out in Professor Djerassi's laboratory independently of DENDRAL support, but of benefit to DENDRAL (62). Other examples will be found in the Bibliography (Refs. 58-70).

3. Existing Capabilities

We have worked to develop distinctive capabilities for molecular structure elucidation, bringing together a high quality HRMS system and AI programs applied to biomolecular characterization. The feasibility of our analytical approach has been demonstrated in several problem areas, based upon the development both of a MS system and a general set of computer programs for use in new areas.

The principal capabilities are summarized below. These are now in being and were developed primarily under NIH funding to this project, with additional support supplied by ARPA and NASA in specific areas. (These agencies have reduced funding levels for this work because overall cutbacks have forced NASA to cut out this area of research despite their interest and ARPA is chartered to provide funds for frontier computer science research but not for applications. Thus the NIH is the principal of support for future development of applications programs in the interdisciplinary area of artificial intelligence/health related chemical problems.)

a. HRMS System and Coupled GC/LRMS System.

We have coupled the NIH-supported Varian-MAT 711 High Resolution Mass Spectrometer with a Hewlett Packard Gas Chromatograph and demonstrated its utility for GC/LRMS analysis of such difficult analytic problems as the free sterols (i.e., not derivatized) isolated from marine and other sources. Advanced data reduction techniques for this instrument were written for use with the ACME computer system (360/50) and now exist in Stanford's new 370/158 which continues to support the PL/ACME language. GC/HRMS scans on extracts from urine and amniotic fluid demonstrated this system's capability to provide high quality mass measurements on complex mixtures obtained from biological sources. An example of one GC/HRMS run on the amino acid fraction of amniotic fluid is presented below (Sec. III.D).

b. DENDRAL Structure Generator (Refs. 1-6,14,31,37,38,40,41)
The DENDRAL Structure Generator program accomplishes exhaustive and irredundant generation of isomers, with and without rings. This program guarantees consideration of every candidate structure - either implicitly, as when whole classes of structures are forbidden, or explicitly, as when individual compounds in a class are specified. It corresponds to the "legal move generator" of computerized chess playing and other heuristic programs.

c. DENDRAL Planner (Refs. 25,28,33)
We have written a very general set of computer programs for determining structural features from analytical data in well-defined areas. Such general planning programs have been written for low and high resolution MS, interpreted proton NMR spectroscopy and 13CMR data.

d. INTSUM (Refs. 26,29,34,35)
INTSUM is a computer program that aids in finding interpretive rules for MS. The program interprets a large collection of MS data according to criteria specified by the investigator. Then it summarizes the data to show which of the possible interpretations seem most plausible.

e. RULEGEN (Refs. 26,35)
RULEGEN is the current rule generation program that suggests various rules of interpretation for the MS data summarized by INTSUM. Although not finished, the program can provide useful assistance in practical theory formation.

f. Ancillary Techniques
1. The MS facility provides other types of experiments in MS, including ultra-high resolution measurements (masses determined via peak matching), defocussed metastable ion determinations (Barber-Elliott technique) and low ionizing voltage experiments. These data are utilized by both scientists and programs where appropriate.
2. Additional computer programs provide added problem-solving assistance.
   a. Predictor program for predicting major features of mass spectra.
   b. Programs for drawing and displaying chemical structures.
   c. Subroutines developed in conjunction with or existing as parts of the Structure Generator for problems of partitioning, construction of vertex-graphs, and constructive graph labelling. These can be applied to answer certain questions of isomerism which do not require the complete generator. For example, the labelling algorithm can list all structures resulting from substituting sites of a carbocyclic skeleton with stated numbers of different functional groups.

g. Other Spectroscopic Techniques
Available to us are the facilities of Professor Djerassi's laboratory for work requiring additional spectroscopic data. Also available on a fee for service basis are extensive spectroscopic facilities (NMR, I.R., and U.V.) of the chemistry department. These would be utilized for collecting additional data on particular structure problems and gathering data on known compounds (particularly in the area of 13CMR) as the AI programs become knowledgable about other spectroscopic information.

h. Chemical Facilities

The staff and facilities of the chemistry department represent substantial synthesis capabilities and general chemical know-how. This resource can be called upon to provide assistance in synthesis of model or labelled compounds, derivatization of mixtures, and so forth. For example, a graduate student in chemistry is presently engaged in thesis research dealing with the laboratory synthesis of a new estrogen metabolite strongly suspected to be a component of certain pregnancy urines. The previously proposed structure of this compound was one of the candidate structures inferred by the planner in a study of estrogen mixtures (11-dehydroestradiol-17-alpha, ref. 33).

4. User Community

Economic utilization of existing and proposed facilities can be realized by sharing them with a community of users. Lacking supplementary funds that would be needed for a comprehensive, major service facility, this community will include the following groups, but will be informally available to others.

A. Stanford Community

i) Stanford Chemistry Department (except for Hodgson, all are heavily supported by the NIH in their research efforts) Letters of interest are attached to the proposal in Appendix A.

Prof. C. Djerassi - Steroids, marine sterols
Prof. W. Johnson - steroids
Prof. E. Van Tamelen - steroids, triterpenoids, other natural products
Prof. H. Mosher - natural products (e.g., marine toxins)
Prof. K. Hodgson - biological ligands, ligand-metal complexes
Prof. J. Collman - cytochrome P450 models

ii) Stanford Medical School Collaborators

The following research projects in the Stanford Biomedical Community will furnish samples for mass spectrometric analysis under the present proposal. Attached to this proposal (Appendix A) are copies of the letters of interest in the proposed facility received from the principal investigators of these grants.

Dr. James R. Trudell, Department of Anesthesia, Stanford University School of Medicine. Drug metabolite identification in humans.

Dr. Irene S. Forrest, Biomedical Research Laboratory, Veterans Administration Hospital, Palo Alto. Drug metabolite identification in humans.

Dr. I. Rabinowitz and D.I. Wilkinson, Department of Dermatology, Stanford University School of Medicine. Prostaglandins.

Prof. Eugene D. Robin, Department of Respiratory Medicine, Stanford University School of Medicine, Ratio of NAD+/NADH in cells by measuring ratio of oxidized to reduced redox pairs.

Dr. Leo E. Hollister, Veterans Administration

Hospital/Department of Medicine, Stanford University School of Medicine. Metabolism of Marihuana.

Dr. Hiram H. Sera, Pharmacy Department, Stanford University Hospital. Drug Identification.

Dr. Sumner M. Kalman, Department of Pharmacology, Stanford University School of Medicine. Drug and drug metabolite identification.

Dr. Jack Barchas, Department of Psychiatry, Stanford University School of Medicine. Neurotransmitters and related compounds in man.

Dr. Keith A. Kvenvolden, Chemical Evolution Branch, NASA Ames Research Center, Mountain View, Calif. Amino acids, acids in geochemical samples, structure of products formed from electrical discharges in gas mixtures.

Dr. William R. Fair, Department of Urology, Stanford University School of Medicine. Identification of the prostatic antibacterial factor; polyamines (putrescine, spermine, spermidine) in body fluids of patients with prostatic carcinoma.

Besides the user projects just summarized, other major prospects are in sight. At the time of writing, the chair of pharmacology is vacant. Conversations with the leading candidate have indicated a deep-seated interest in GC/HRMS as the principal analytical tool for broad ranging studies of drug metabolism in man.

B. Extramural Users

The development of the techniques of ORD, MS and MCD at Stanford has been paralleled with extensive sharing of these resources nation- and world-wide in collaborative research efforts, without any additional funding. Rather than provide routine service, experience has shown that discretionary selection of problems results in better utilization of our people and instrumentation resources. We would extend this provision of services including available computer programs, to a limited number of extramural users. Note, for example, our successful collaboration with Professor Adlercreutz, Meilahti Hospital, University of Helsinki, on the identification of estrogens from body fluids utilizing the AI planning program (ref. 33).

C. Relationship to AIM-SUMEX and the Genetics Research Center

The present application is strengthened by two research projects related to, but not overlapping, the proposed research of this grant.

1) AIM-SUMEX (NIH RR-00785, Oct. 1, 1973, thru July 31, 1978, Principal Investigator, J. Lederberg). This is a resource grant to establish a national facility for applications of artificial intelligence in medicine (AIM). Our own use of this facility will include SUMEX PDP-10 computer time and file storage necessary to run the DENDRAL artificial intelligence programs. This support will be furnished without charge to the present proposal. It represents an annual investment of about $100,000 in computer time equivalent value.

The AIM-SUMEX computing facility is shared equally between a national user community (AIM) and a Stanford Medical School community. The DENDRAL research will be supported out of the Stanford portion. The AIM service will be administered under the policy control of a national advisory committee and will be implemented over a national computer network. AIM-SUMEX provides the means for members of the national user community interested in structure elucidation to access the DENDRAL programs.

2) Genetics Research Center (NIH PO1-GM 20832-01 - approved by the NIGMS Council, awaiting funding, Principal Investigator, J. Lederberg). This research proposal is a comprehensive grant which would support interdepartmental research at the Stanford Medical School in Medical Genetics, Pediatrics and other clinical applications. A section of that proposal concerns the use of GC/LRMS for screening body fluids for evidence of inborn errors of metabolism. (This project grew out of the initial DENDRAL grant, one of the research goals of which was the analysis of body fluids using GC/MS). This research on inborn metabolic errors will be conducted jointly in the Stanford Departments of Genetics and Pediatrics using existing equipment (Finnigan 1015 Quadrupole mass spectrometer, Varian Aerograph GC and a PDP-11/20 based data system).

We appreciated the value of GC/HRMS analyses of selected extracts of body fluids (i.e., those containing metabolites not identified by routine GC/LRMS data) when formulating the Genetics Research Center proposal. Accordingly, a small amount of funding was there requested for recording selected GC/HRMS data on the GC/Varian MAT 711 mass spectrometer in the Department of Chemistry. If these funds are awarded, we will negotiate with NIH a suitable elimination of this minor overlap with the present budget.

II. SPECIFIC AIMS

The specific aims enumerated in this section will be pursued in the highly inter-disciplinary manner that has characterized the DENDRAL project from the start of its NIH support. The aims are not disjoint, but interactive and inter-dependent. For example, the power of MS and, potentially, other spectroscopic techniques, can be enhanced by the use of computer programs to perform various aspects of structure elucidation and theory formation. From the standpoint of computer science, one measure of the utility of techniques of artificial intelligence is how well they perform in real-world applications. It is necessary in the development of these programs to have a source of data and informed, involved team-mates able to criticize methods and results. The aims are elaborated in the methods section.

We have attempted to keep the proposal to a readable length. Therefore, some detail has been omitted. However, many details can be found in the biliography and we are prepared to provide additional information during the site visit.

1. Enhance the power of the MS resource.

The existing MS resource, together with computer programs which exist or which are proposed (see Aim 2, below), is capable of solving some of the structure elucidation problems of the user community given computer support for data collection and reduction. We refer specifically to the areas of GC/LRMS and routine, batch HRMS samples. We believe that many of the problems of the user community require more powerful techniques (see Section III). These techniques, specifically GC/HRMS and semi-automatic metastable defocussing, can be provided with a minimum of cost and effort, thus enhancing considerably the capabilities of the resource.

Our first aim is to provide the resource with adequate computer support (replacing the previous ACME system) to enable collection and reduction of mass spectral data including low and high resolution scans and data on defocussed metastable ions.

We propose to develop this computer support in the ways described below. (these aims are written to include the work necessary to implement the extended PDP-11/20 computer system. A description of the rationale for this choice is provided in Section III.A and the specific augmentations in the Budget Justification).

A) Convert existing, proven data acquisition and reduction programs from the PL/ACME language into Fortran, consistent with time-critical assembly language programs for data acquisition and instrument control. These programs will be written in Fortran to enhance compatibility with the computer systems of other users of such packages.

B) Modify these programs, as required, to handle acquisition and reduction of frequent or repetitive HRMS scans with selected instrument performance feedback to the operator, and to take advantage of the expanded capabilities of the extended 11/20 system. Prototype GC/HRMS systems have been developed at Stanford and elsewhere, but this type of facility (in contrast to GC/LRMS)

is not now available to the Stanford community. When this system is developed, service will be available to the Stanford community and research collaborators and, if our resources permit, to any scientist requesting assistance. In many instances this type of collaboration will require far more involvement of convergent interests, efforts and skills than merely running samples on request. We have in mind the chemical and eventually biological interpretation of the analytical data as a matter of joint concern, as appropriate.

We have previously illustrated the advantages of high resolution mass spectral data in the computer analysis of mass spectra (e.g., ref. 28). Also, we have previously shown that the same program can deal with analyses of mixtures without prior separation especially when additional data (e.g., from selected metastable defocussing experiments) were provided (Ref. 33). We wish to use the MS resource and the computer program in further studies of mixtures of compounds which are difficult or impossible to separate by GC. The advent of routine systems for high pressure liquid chromatography have made many of these separations possible, but the liquid chromatograph is not presently interfaced to the MS.

Many of the problems of the user community require analysis of complex mixtures which are amenable to treatment by GC/MS techniques. We feel that where sample quantities permit, acquisition of GC/HRMS data is highly desirable. These data can be provided by the resource supplemented with computer support (above).

We propose to continue tests of the GC/MS combination, operating under moderately high mass resolutions (5000-10000), to define in detail the optimum operating conditions of the GC/HRMS combination. This will provide the necessary information on maximum practical sensitivity to be expected. This information can then be used in collaboration with the user community for sample preparation.

The GC/HRMS system would normally be operated at reduced mass spectrometer resolutions to maximize sensitivity. We have existing multiplet resolution programs to increase the resolving power of the MS. We propose to provide the multiplet resolution program with heuristic guidance based on compositional variations inferred from molecular ions or other singlet peaks. For example, a resolving power of 10,000 is barely sufficient to resolve ions which differ by CH2 vs. N (delta m = 0.012) for ions of about mass 100. Although it will resolve CH4 vs. O doublets (delta m = 0.037) at this mass, it will not resolve closer doublets such as C3N vs. H2O3 (delta m = 0.003). We can provide exhaustive tabulations of multiplets by mass separations (based on ref. 30) which can be used by the multiplet resolution program.

We have previously indicated the power of metastable ion information in the operation of our programs for structure elucidation (refs. 28, 33). We have extended one of our programs (the MS predictor program) to propose metastable defocussing experiments in order to avoid collection of unnecessary data (see Aim 2, below). Although we can collect these data (Barber-Elliott technique) manually on our existing Varian MAT-711 MS, this is an

exceedingly wasteful operation, both in terms of sample consumption and time. We propose to implement some automation of collection of these data on metastable ions. We also propose to begin preliminary investigation of alternative modes of metastable ion determination (see Methods (Sec. III), below).

2. Develop performance and theory formation programs to assist in the solution of structure elucidation problems in biomedicine.

Computer programs have already been written for analysis of low and high resolution mass spectra, for generation of acyclic and cyclic molecular structures, for labelling structural skeletons with atoms, for analyzing 13CMR spectra of amines and for interpretation and summary of large volumes of data gathered on model compounds (see Existing Capabilities above, for references). We wish to increase the utility of these programs by providing interactive facilities that allow easier access to them, by increasing their generality and power, and by supplementing them with new reasoning programs.

Performance Programs:

The current structure generator program will be subjected to further detailed tests before using it for structure determination problems.

A new algorithm for generating cyclic skeletons (with no multiple bonds) will be programmed and checked. The algorithm is written and informally proved. A formal proof will be devised as well. This algorithm represents one very powerful approach to the problem of implementation of constraints, as discussed in the following paragraph.

The generating programs will be modified to allow isomer generation within constraints. Different kinds of constraints can be inferred from different kinds of spectroscopic data. We intend to give the program knowledge of a variety of these.

The Planner programs that infer constraints from mass spectrometry data will be broadened to include additional knowledge about the spectral behavior of classes of compounds of relevance to the NIH-sponsored research of the user community. In addition, we will add the capability for utilization of information about chemical isolation procedures (e.g., one expects acidic and neutral compounds in solvent extraction of acidified body fluids) and relative GC retention times (e.g., to admit the possibility of homologous series).

We propose to implement a more general method for inferring the identity of the molecular ion whether or not this appears explicitly in the spectrum. This information is important for the successful operation of the structure generator and the planner. We want the program to use whatever information is available and not depend, as it currently does, on having knowledge of the structural class together with inference rules for that class.

Interface routines will be written to make it easier for other scientists to use these programs. We have to wait for an

interactive system before starting this: AIM-SUMEX will be ideal. Input/output routines will be crucial to easy use of the system. However, we also want to give users the facility to understand the system's reasoning steps so they can take advantage of it.

In addition to making the computer programs available through AIM-SUMEX, we would like to translate parts of the LISP code into another language - for reasons of both efficiency and exportability. We have talked with computer professionals at IBM Research Center about using the APL language. FORTRAN, ALGOL and PL/1 are other languages whose merits for our purposes we will explore.

We wish to continue a low-level of effort on computer programs that interpret other kinds of spectroscopic data. Planning programs similar to the MS Planner could be written for automatic analysis of data from other spectroscopic techniques(e.g., IR, UV), as we have illustrated for 13CMR (ref. 39).

The structure generator's view of chemical structure is topological and is presently unconstrained by bond lengths and angles. Because stereochemical considerations are frequently important in structure elucidation, we propose to begin consideration of stereochemistry in the structure generation and evaluation processes.

A program with detailed knowledge about information obtainable from various spectroscopic techniques could be written to examine a list of candidate solutions and propose experiments necessary and sufficient to distinguish among them. The program would represent an extended Predictor (e.g., ref. 27). We have a first version of a program that suggests "crucial" metastable peaks to be sought in order to distinguish among candidate structures. Work on this program will continue at a low level of activity, possibly expanding into areas other than MS. One topic we will continue to pursue is our collaborative effort with Dr. Gilda Loew, Genetics Department, on the potential application of molecular orbital theory to prediction of mass spectra (ref. 71).

Theory Formation Programs:

The rule formation program (RULEGEN) will be extended so that it can search a larger space of rules. Present a priori constraints on the rule generation give us a search reduction from tens of millions to a thousand possible rules. Even though search heuristics now allow efficient search of these possibilities, we want to be able to deal with much larger spaces efficiently, as when the number of primitive predicates is drastically increased.

The RULEGEN program will be modified so that complex fragmentation and rearrangement processes are manipulated nearly as easily as simple fragmentations. The program currently finds fragmentation rules involving one or two bonds, possibly followed by hydrogen migration. In the case of cyclic systems such as estrogens, however, the program must be able to work with sets of three or more bonds in some cleavages.

Interactive programs will be provided on AIM-SUMEX for the investigator to query the rule generation program. For example,

many questions now arise about the program steps by which the program infers the rules it suggests as explanations of the regularities. Why, for example, was some particular rule not considered plausible?

New data will have to be selected in order to test the rules and to differentiate among competing rules. We will write a program that suggests new experiments (i.e., new data to obtain), depending on the nature of the existing rules.

The test phase of the theory formation program will be written as an evaluation function of each rule against new data. Insofar as any new experiments are "crucial" experiments, the evaluation function may merely reject a proposed rule. Mostly, however, rules will have to be evaluated against new data along many dimensions: frequency, strength of evidence, uniqueness, simplicity, and the like.

We wish to experiment with the whole theory formation program to determine the critical aspects of our design. For example, (1) how sensitive is the program to discrepancies, inconsistencies and errors in the data? (2) how well can the program find rules within a slightly different model of chemistry? (3) how well can the program perform with one pass through the data, or several passes? and (4) how critical are the principles of theory formation?

3. Apply the structure elucidation techniques - both instrumentation and computer programs - to biomedically relevant compounds.

Our own interests are in elucidating the structures of, and understanding the MS of, marine sterols, hormonal steroids, and compounds isolated from human body fluids that can be associated with genetic disorders (from research in the GRC). In addition, we will be working closely with members of the Stanford Medical School and Chemistry Department - in particular those mentioned above (Section I.B.4) - on their structure elucidation problems in which MS will be used. Although most users expect to require HRMS and GC/HRMS data, some of their problems will be attacked utilizing GC/LRMS techniques and library search through (usually) restricted libraries of mass spectral data. We propose to investigate some extensions to the technique of library search (see Methods) to complement our existing and planned DENDRAL programs. We plan to continue our exchange of mass spectral data and library search information as we have previously done with Dr. S. Markey (University of Colorado Medical School) and Dr. F. W. McLafferty (Cornell University).

As in the past, attention to new biomedical research problems will lead to increased capabilities in the computer programs. We require close communication with the people engaged in the research so that the programs actually assist the researcher while increasing in power. Collaborative proposals have come out of such past DENDRAL sponsored work, for example, large portions of the GRC proposal and a proposal for 13CMR research.

We envision the interaction and collaboration with the user community to involve the following:

a) In all cases, we plan close cooperation with the users in all aspects of the problem. Although the basic isolation procedures are the problem of each investigator, his knowledge of the available facilities and their limitations can be an important aid to sample preparation and analysis of the results. This is particularly true for collaborators who are unfamiliar with the techniques of HRMS (e.g., sample size and resolving power necessary to separate the mass doublets that can be realistically expected in different contexts).

b) The needs of the user community will be varied. Drs. Duffield and Smith will, in collaboration with the users, determine the kinds of MS experiments which will be most useful, considering sample complexity, stability, quantity, and so forth. We wish to utilize fully the existing resource and our proposed extensions, bringing to bear on a problem any technique which is appropriate and can be provided. This will include the full scope of available experimental techniques in MS (LRMS, HRMS, GC/LRMS, GC/HRMS, metastable defocussing, ultra-high resolution mass measurements) and available computer programs (see below).

c) Many problems will be amenable to treatment by computer programs which exist or which will be developed, for example, structural isomer problems or HRMS interpretation on compounds in a well-understood class. We will take the responsibility for utilizing these programs where appropriate to assist in structure elucidation problems. We will instruct members of the community in use of the programs when programs are used routinely by collaborators.

III. METHODS

Molecular structure elucidation entails the intelligent and patient application of a large body of knowledge to each specific problem. The importance and relative difficulty of the problem impel us to seek the powerful assistance of computer programs to help chemists in their analyses. It is unlikely that such programs will ever replace chemists, especially because computer programs are readily written only to focus on rather narrow aspects of problems. Nevertheless, our past research is reasonably forwarded as a demonstration of the computer's ability to assist in practical biomolecular characterization although this was a spinoff from theoretically oriented research.

In order to meet the major objectives of this proposal we will focus our attention primarily on structure elucidation of biomedically important compounds through MS and AI. However, many of the computer programs can already use information from other analytical techniques. So we want to be able to think of structure elucidation in the context of an ensemble of analytical capabilities.

A. Enhancing the Power of the Mass Spectrometry Resource

We have developed a significant resource consisting of instrumentation (the Varian MAT-711 and ancillary equipment) and computer programs for instrument evaluation, data acquisition and reduction. Routine reduction of high resolution mass spectra to elemental compositions and ion abundances without human intervention provides the capability for efficient handling of large volumes of high resolution mass spectra (such as will result from GC/HRMS runs). The development of the GC and of the GC/MS combination is in the excellent hands of Ms. Annemarie Wegmann, who is responsible for operation of the complete system. We now have more than two years of operational experience with the MS, the GC and related equipment under a wide variety of experimental conditions.

None of the resource-related research discussed in this proposal can be carried out without significant quantities of mass spectral data. The existence and extensions of the MS resource, the development of computer techniques and the applications to biomedical problems demand an efficient mechanism for acquisition and reduction of MS data, and eventual transmission of the data to the SUMEX resource. Thus, operation of the MS requires substantial computer support to deal with the large volumes of data produced by the system at high data rates. We feel that a properly configured system of hardware and software should provide, at a minimum, the following capabilities:

1) Detailed evaluation of the condition and performance of the MS prior to recording data on valuable samples, with feedback to the operator.

2) A coordinated system of hardware and software for signal conditioning, peak detection and peak analysis.

3) Data reduction techniques based on a computed (not theoretical) model of the MS, including peak shapes, mass/time function, and resolving power as a function of mass.

4) Peak profile analysis for multiplet detection and resolution.

5) Computer control of scan rates, clock rates and optimum analog and digital filtering parameters.

6) Some on-line feedback, to the operator, assessing the performance of the system during an experiment.

7) The system must deal with frequent or repetitive HRMS scans, requiring the capability for rapid storage and analysis of large volumes of data.

Previous support of our research by the NIH and NASA has given us a firm foundation of programs and experience. We have, up until the termination of the ACME computer facility (July 31, 1973), demonstrated capabilities 1-5 above. We were precluded from pursuing capabilities 6 and 7 due to the configuration of the ACME facility.

The demise of the ACME computer facility and the subsequent incorporation of the PL/ACME language into a new IBM 370/158 facility under Stanford auspices has forced a reevaluation of the means for providing HRMS laboratory computing support. We had previously depended exclusively on ACME for data reduction processing. The ACME transition poses both technical and fiscal decisions in that the real-time support capabilities of the new facility will be different from ACME's and the fee for service basis of the facility requires an explicit budget allocation for its use. Previously we had received ACME computing support without charge as part of the core research effort. Since we were thus required to revise our computing plans, we have explored a number of options for near-term as well as longer term solutions.

As outlined in the attached annual report, we have chosen an interim approach (through the end of the current grant, 4/30/74) which minimizes near-term costs, including hardware and software conversions as well as operating expenses. This approach entails connecting the MAT-711 spectrometer to the 370/158 computer through an IBM 1800 interface. It allows use of the existing PL/ACME programs but will have real-time response limitations at least as severe as ACME had (which is inadequate for either GC/LRMS or GC/HRMS). Our existing computing budget provides for only a very low level of instrument utilization in this mode.

For a longer term solution these constraints are unacceptable. Current estimates are that continued use of the inadequate 1800-370/158 connection and PL/ACME interactive programs under full instrument productivity would cost up to $4,100 per month. Three alternatives have been investigated for improving technical performance and reducing cost. This review has resulted in our current proposal to augment the existing mini-computer system (PDP-11/20) with local storage and arithmetic capabilities. This stand-alone system would not support real-time, on-line data reduction but would allow routine data acquisition and instrument performance evaluation, followed by off-line data reduction.

Alternatives considered include:

1) Modified 370/158 Connection

We discussed with personnel in the Stanford Center for Information Processing (SCIP) various approaches for improving 370/158 service. Detailed planning is still under way within SCIP in regard to real-time support and future pricing policies. Thus the following conclusions are tentative. It appears that the long-term cost would be prohibitive to continue real-time data acquisition by the 370/158. Instead, a store-and-forward system was proposed. This would entail an augmentation of the existing PDP-11/20 front-end mini-computer with memory, disk, tape, and a new interface to the 370/158, totaling about $28,000. This approach is workable, if limited near real-time instrument performance evaluations could be made to assure satisfactory instrument setup and data acquisition. It was recommended that the existing software be converted from PL/ACME to a more efficient language (such as FORTRAN) to reduce operating costs. This would require approximately 4-6 man-months of effort. The resulting decrease in operating costs could not be estimated in time for this proposal because the new SCIP pricing policies are not formulated and inadequate 370/158 system analysis tools are operational to evaluate our benchmarks in terms of detailed resource consumption. We have therefore budgeted an approach based on the remaining two options with the understanding that we will reconsider the SCIP option before proceeding with an implementation should this proposal be funded.

2) SUMEX

The recently approved AIM-SUMEX PDP-10 facility will provide necessary computing support for the development and use of DENDRAL AI programs. The MS laboratory produces data which these programs analyze and thus has a close relationship to the AI research. The SUMEX computer could help in the off-line reduction of instrument data, particularly during the early stages of the project when the machine load will be relatively light (20-25%). The present programs would require conversion from PL/ACME as in option (1), which would take 4-6 man-months. Such computing may use from 15-30 minutes of CPU time per day, depending on the amount of GC/MS work. While this approach saves operating computing costs, the front-end PDP-11/20 would require augmentation as in option (1) ($28,000) to allow store-and-forward operation with subsequent off-line data reduction on SUMEX. This is needed because SUMEX is not configured to allow real-time acquisition of the volume of data anticipated. This approach, while the least costly, would entail a measurable use of the PDP-10 resources which we feel are better reserved for the intended AIM-SUMEX applications. In addition, because of the priorities anticipated for allocation of SUMEX to AI research, particularly as loading increases, scheduling may be required which will constrain the MS laboratory operation. For these reasons, we feel a better, even though slightly more expensive, approach is a stand-alone PDP-11/20 data reduction system.

3) Stand-Alone PDP-11/20

The augmentations of the existing front-end PDP-11/20 required for store-and-forward operation in conjunction with the 370/158 or with SUMEX, come close to meeting the needs of a stand-alone data system. In addition to the memory, disk and tape, an augmented arithmetic capability is needed to allow rapid floating point calculations. A special device for this purpose costs about $7,500. The SUMEX interface can be less sophisticated in this case, however, accounting for the much lower data volume after reduction, so that the total cost of the stand-alone system would be $34,000. As with the other options, conversion of the present programs would be required.

This approach, while slightly more expensive, has the advantages of off-loading all data logging and reduction functions from SUMEX and affords an adequate capacity for non-real-time, stand-alone data reduction on the PDP-11/20. It furthermore allows more freedom and responsiveness in the operation of the MS laboratory since data collection or reduction can be scheduled without worrying about the impact on AIM-SUMEX users. We therefore propose and have budgeted an augmentation of our existing mini-computer system as a stand-alone data reduction facility.

The biomedical community (see User Community, Sec. I.B.4 above) desiring access to our facilities for structure elucidation have a variety of problems, some of which can be solved by existing instrumentation and computer techniques, as noted above. However, many problems consist of complex mixtures of compounds where analysis by conventional GC/LRMS does not lead to unambiguous solutions, and separation of components on a preparative scale for other spectroscopic analysis is difficult (e.g., see marine sterols, section D, below). These problems are amenable to attack by a system comprised of a GC/HRMS combination, the GC providing separation, coupled with the MS operating at high resolution to provide elemental compositions. Thus, upgrading of our current system so that GC/HRMS data can be provided on a routine basis is a desirable, and we believe necessary, step to solve many of these problems.

We propose to continue the development of the GC/HRMS system while maintaining existing capabilities of routine HRMS analysis and GC/MS where this efficiently responds to local needs. Many members of the user community will require in addition to GC/HRMS, HRMS analysis of relatively pure compounds or mixtures of small numbers of compounds. We will provide this capability on an interim basis, using Stanford's IBM 370/158 system while the PDP 11/20 system is being upgraded.

We were able, using the ACME computer facility, to start evaluating the operation of a GC/MS system at high mass resolutions. These experiments were hampered somewhat by the limitations of the computer system used to acquire the data (only occasional, single scans were possible); they were necessarily discontinued (as well as all HRMS operation!) upon the termination of ACME. We do have, however, some benchmark figures for the performance of the proposed system. Mixtures of fatty acid esters (e.g., methyl palmitate and methyl stearate) gave good quality

mass measurements (+-10 ppm) over a dynamic range of 100:1 for sample sizes of the order of 0.5-1.0 micrograms/component during 10 sec/decade in mass scans (resolving powers 5,000-8,000).

We are haltingly continuing our evaluation of the GC/HRMS system even without a data system, making measurements on individual ions of the mass standard and known materials in the GC effluent. These data can be approximately translated into expectations during dynamic scanning. We have performed an extensive series of measurements on both methyl stearate and cholesterol (not derivatized), the latter compound being more representative of our current research problems. These measurements tend to confirm the preliminary data described above. Firmer data will be available subsequent to the submission of this proposal.

We propose to operate our existing GC/MS system under high resolution conditions aiming toward optimization of resolving powers, scan rates and GC and molecular separator operating conditions to determine the maximum usable sensitivity of the system.

We recognize that the ultimate sensitivity will not approach that attainable by photographic methods of recording; we feel that the ability for on-line operation and evaluation of the operating conditions of the MS partially offsets the sensitivity disadvantages. We realize that some structure elucidation problems will not be amenable to study because of the sensitivity limitations; we feel, however, that many problems of interest to the User Community can be studied effectively with this performance capability. Rather than propose a research program to increase the sensitivity of high resolution mass spectrometers (e.g., McLafferty, et.al., Anal. Chem., 44, 2282 (1972), dynamic rescanning of peaks; Jet Propulsion Laboratory - chemical multiplier emission/detector arrays, private communication to T. Rindfleisch), we propose to identify our limitations and, with our collaborators, use discretion in selecting and preparing samples.

Further accelerations of technical capability to meet the state of the art in sensitivity will require investments in hardware that can be better justified at a later stage of a successful facility program . Meanwhile, other laboratories can be expected to make significant contributions to this important problem. Practical regard for budget limitations is the main reason we do not press this issue ourselves at the present time.

Significant improvements in sensitivity (with only small decreases in mass measurement accuracy) can be achieved by operating the MS at reduced resolving powers coupled with intelligent analysis of the resulting data to detect and resolve the potentially greater number of overlapping peak envelopes. This proposal is not entirely new (e.g., see Smith, et.al., Anal. Chem., 43, 1796 (1971); Burlingame, et. al., in "Computers in Analytical Chemistry," C.H. Orr and J.A. Norris, Ed., Progress in Analytical Chemistry, Vol. 4, Plenum Press, New York, N.Y., 1970, Chap. III). We can, however, significantly extend these earlier techniques by utilization of our multiplet resolution algorithm. This algorithm embodied in a computer program, has been shown to increase the effective resolving power of the MS up to a factor of three. It bases its operation on a dynamic model of peak shape computed directly from the data. For computational efficiency and to avoid

spurious information, this algorithm would be best implemented as a post-processor, basing its search for multiplets on the results of prior elemental composition determination.

The ability to detect and analyze for unresolved peaks is mediated by consideration of the mass measurement accuracy of an MS system. These systems are capable of determining peak positions (and thus masses) to a small fraction of the peak width. The high accuracy of such measurements (+- 2-10 ppm) can, in fact, be utilized to detect and "resolve" multiplets in instances where the unresolved species are known precisely (see Burlingame, et al., ref. above, for CH vs. 13C doublet detection and resolution).

For instances where the heteroatom content of a molecule is known or where the possibilities are reduced severely by chemical, spectroscopic and mass measurement heuristics, there may be a range of possible overlapping ions resulting from fragmentation of the molecule. These potential overlaps may be computed and then used (in combination with the known resolving power and mass measurement accuracy of the MS and the measured mass of the peak, assuming it was comprised of only one type of ion) to direct the multiplet resolution program.

As an example, we have computed the possible mass doublets for various ranges of compositions (Lederberg, et al., to be published). A sample table for C, N, O =<4 is appended (Table 1). Only 28 of the 364 possibilities are shown, namely those whose mass difference (e) <.05 mass units. *

Of these 28, 13 show e>.03 and would be fairly easy to resolve, requiring 1/5000 resolution at MW=150.

At the other extreme, 5 doublets show e<.01 (CN4 vs. H4O4; C2H2O vs. N3; C2N2 vs H4O3; C3N vs H2O3; and C4 vs H2NO2) which would demand special treatment for resolution.

The 10 doublets for which .01 =< e =< .03 pose the interesting challenges for tradeoff of resolution vs. sensitivity in the context of given problems. For example, if N is absent, the only ambiguities are C3 vs. H4O2 (e = -.02) and C4 vs O3 (e = .015).

Much as we would wish always to have unambiguous empirical formulas for all ions, HRMS remains a valuable tool despite these limitations. As shown by these examples, even moderate resolution reduces the number of candidates to a manageably small number of alternatives. Contextual and interval data (within the spectrum) can be used to trim these further at two levels: (a) pooling of peak statistics to sharpen decision probabilities on the presence of heteroatoms -- the fragments are subsets of the molecule and (b) the assemblage of candidate solutions under each of the alternative formulae. Manifestly, computer processing can sort out branches of decision trees that would soon exhaust human patience.

These heuristics are built into the DENDRAL programs (solutions based on fragmentation theory), but are also applicable to table look-up approaches.

We (ref. 28,33), and others (e.g., H.-K, Wipf, et. al., J. Amer. Chem. Soc., 95, 3369 (1973)) have illustrated the importance of

metastable ion determinations in automated structure elucidation based on MS data. Data on metastable ions must be judiciously selected because of the time and sample normally required to perform the measurements. Our programs are now capable of precise specification of those experiments necessary and sufficient to distinguish among a set of candidate structures. We seek more efficient ways of acquiring these selected instrumental data. This can be accomplished with minimal cost by developing the hardware and software necessary to perform (defocussed) metastable scans and calculate the data. Much of the hardware, except an accurate sensor for accelerating voltage, already exists. We have had considerable experience in peak detection on the software side; the calculations to determine transitions are simple. It is assumed that the operator would manually adjust the instrument to the desired "daughter" mass prior to initiation of the scan of metastable origins ("parents") of this daughter.

The recent availability of reversed-geometry instruments has provided new methods of metastable defocussing (e.g., Beynon, et.al., Anal. Chem., 45 (12), 1023A (1973)). We have illustrated the power of these techniques in mixture analysis (ref. 69). No "normal" geometry instrument is equipped to perform these measurements to determine all the daughters of a given parent, information which is frequently more useful than the converse. This information can be obtained, in principle, by synchronous variation of two of the three fields (magnetic/ accelerating/ electrostatic deflection) in a very accurate way. We would like to explore this possibility because we feel that this technique, if feasible, would represent a significant upgrading of the many standard geometry, double-focussing instruments in existence.

B. Computer Assisted Structure Elucidation

As mentioned above, some existing programs can be used immediately for structure elucidation problems using MS data. The programs have been described in detail elsewhere and are mentioned in the section on existing capabilities (Sec. I.B.E, above). The Planner's performance, for example, is excellent precisely in the areas where MS, by itself, is capable of definitive structure analysis. The general intellectual flexibility of the human chemist is beyond the reach of plausible programs. On the other hand, where the history of a sample is known, so as to restrict the potential classes of compounds and for classes where the rules of MS fragmentation are well understood, the program's performance matches that of trained mass spectroscopists, the program also offers some advantages in its exhaustive and rapid analysis of the data. Many structure elucidation problems of the user community fit into this category and existing resources can fulfill these needs.

Whether man- or computer-implemented, MS cannot solve all structure elucidation problems, however. In such cases, recourse is to other spectroscopic techniques if sample size permits. As described in the introductory section, diverse information is pieced together to achieve a solution. Interactive computer programs can assist in segments of this procedure, with the advantages of exhaustive evaluation of the data and the molecular structures suggested by these data.

In our own and in planned collaborative work, we will call upon the extensive facilities of the chemistry department for acquisition of additional spectroscopic data. These services are financed by fees, paid from existing research grants of the user community. There are sufficient documented examples of structure elucidation problems to obviate the requirement for extensive use of these additional facilities in development of the programs. On the other hand, the intensive pursuit of mechanized "intelligence" in the domain of MS requires more than availability of public MS data . It requires the collaboration of skilled chemists actively engaged in practical MS research and, at the same time, committed to the exploration of innovations in the application of AI to the solution of the problems

As in the past, we will develop the computer programs through close collaboration among Drs. Duffield and Smith (and other members of their groups) and the program designers and programmers. For us, this means daily consultation for discussion of strategy, extensions to the program, and solutions to new problems. In particular, we propose to continue software development (on the AIM-SUMEX facility) as folllows:

1) The recently completed structure generating algorithm will be the core of our efforts to assist in structure elucidation. The structure generator can guarantee that the correct solution is somewhere in the list of possibilities. Additional programs, such as the Planner allow us to avoid exhaustive generation in practice. Some parts of the cyclic structure generator program have not been extensively tested yet, and these tests will be the first task to complete.

2) The structure elucidation task is strongly directed toward rejection of whole categories (e.g., compound classes) of solutions as quickly as possible by using as much knowledge about the chemical history or characteristics of a sample as is available. Details of spectroscopic data then define the molecular framework more precisely. Each step in this procedure represents the application of constraints on the set of possible solutions. Computational efficiency demands that these constraints be applied early in the generation process when the structure generator is utilized.

We have made some effort to examine the kinds of constraints used by scientists engaged in structure elucidation. We have begun designing strategies so that these constraints can be brought to bear on the structure generator. Some of these strategies involve minor changes to the existing program; others require significant extensions of existing generating functions. One approach which seems particularly attractive to us is presently under development. This approach will utilize the existing structure generator, with some modifications, to generate a dictionary of cyclic skeletons up to those containing a maximum of twelve tertiary vertices. The dictionary will be a complete, irredundant list of ring systems which contain no multiple bonds and no cut-edges (acyclic parts). This dictionary will be organized and keyed so that many constraints can be implemented easily. The dictionary will allow exhaustive specification of ring systems with double bonds and/or aromaticity. The rings themselves can be labelled with heteroatoms to generate heterocyclic ring systems,

or with acyclic radicals to generate substituted ring systems. The existence of the dictionary will lead to greater computational efficiency as it needs to be generated only once, and specific configurations of rings (numbers, sizes, fusions) can be pulled immediately from the dictionary.

We propose to continue these investigations so that a reasonable variety of constraints can be recognized and utilized effectively by a computer program. This represents the first step toward increasing the chemical knowledge of a program which views molecular structures and their manipulation as mathematical entities and transforms.

3) Present, effective use of the structure generator or its subroutines for special problems requires a detailed knowledge of the program. We propose to develop an interface between users and the program to remove this requirement. The interface would contain elements of structure input and display routines and a simple language for application of constraints. Portions of these elements are available from other workers (e.g., Richard Feldman, NIH) and we would draw on these sources whenever possible.

4) We propose that initial efforts will be directed toward a system where the scientist examines his own data and inputs his findings (in terms of allowed and disallowed structural features) to the program as constraints. The generator would then provide a list of possible solutions to be evaluated, followed by iteration on this procedure.

5) Many structure elucidation problems can be characterized as assembly of sub-structures inferred from spectroscopic data into complete molecular structures. Although there are two instances in the literature describing programs with the capability to solve this problem (see S. Sasaki, "Determination of Organic Structures by Physical Methods, Vol. 5," F.C. Nachod and J.J. Zucherman, Ed., Academic Press, New York and London, 1973, p. 285; M.E. Munk, C.S. Sandano, R.L. McLean, and T.H. Haskel, J. Amer. Chem. Soc., 89, 4158 (1967)), we do not feel these approaches fulfill the requirements for generating complete lists of structures and avoiding duplicate structures. We have some strategies to solve this problem, thus extending the scope of the generator while tying it more closely to the methods used by chemists engaged in structure elucidation. Our existing structure generator has this capability; as long as the sub-structures are connected only by a single bond, no new rings are formed.

6) We wish to implement general routines for finding molecular ions from spectroscopic data in order to improve the general power of the Planning program. The current Planning program depends on having some metastable ion information with HRMS data, together with knowledge of the structural class with special rules for the class. We will incorporate strategies suggested by Biemann (K. Biemann and W.J. McMurray, Tet. Lett., 647 (1965)) and McLafferty (R. Venkataraghavan, F. W. McLafferty, and G. E. Van Lear, Org. Mass Spectrom., 2, 1 (1969)) for finding molecular ions, but also give the program the flexibility to use class-specific information when available. The procedure will be to use these kinds of information within a general heuristic search paradigm.

7) The section on aims indicated some longer-term directions

which might be pursued. Of these, we feel that the incorporation of three-dimensional information into the program is perhaps most important (e.g., representation of three dimensional information, molecular modelling including steric factors). Lederberg has previously discussed ways (Ref. 1) in which three dimensional information can be considered in the generation and representation of molecular structures. More recently, the work of Wipke (J. Amer. Chem. Soc, in press; personal communication) in connection with computer assisted organic synthesis has provided important results which we would attempt to utilize to avoid unnecessary duplication of effort. We plan to collaborate with Dr. G. Loew (Stanford Genetics Dept.) to utilize her available programs on molecular orbital methods to determine local minima for conformations.

Another longer term goal which we feel is both interesting and important is the use of an extended Predictor (which we have previously described in the context of MS) to assist in distinguishing among potential solutions to a structure elucidation problem. We have recently carried out some extensions to the existing Predictor by incorporating the ability to suggest metastable defocussing experiments. Further extensions to include knowledge about other spectroscopic techniques and the information which can be elicited from these techniques are clearly feasible and could be a powerful extension to our computer assistance efforts.

C. Theory Formation

One important aim of this project is to improve the existing theory formation capabilities and thus provide more assistance to scientists investigating regularities within classes of compounds. This is a theory formation task at a very pragmatic level. The MS theory that the program attempts to find is of the same form as the one practicing mass spectroscopists use for structure elucidation. Thus, resulting pieces of theory are extensions to both the scientists' theory and the computer's theory of the discipline. To improve this program we need to complete the Plan-Generate-Test program that has been started (as described in the appended annual report) and tune it over many test cases. We also wish to make the programs interactive and easy to use so that they are more readily accessible. This can be done when the programs are transferred to the AIM-SUMEX facility.

We plan to apply the theory formation program to two different kinds of data: (a) the data collected in the interest of understanding the mass spectrometry of a particular class of compounds, as was done for estrogenic steroids, and (b) collections of diverse data that may provide some insight into more general fragmentation mechanisms. For example, we hope to find general rules analogous to the alpha-cleavage rule or the stability of aromatic rings.

The INTSUM program mentioned in Section (I) is the planning phase of the theory formation program. It currently runs in batch mode on Stanford's 360/67 computer. We wish to add an interactive monitor to INTSUM to give an investigator the ability to set up his own conditions for interpreting the mass spectra and to control the type of summary he wishes to see. For example, if he

is interested in the allowable hydrogen transfers associated with one specific process the program could be instructed to produce a very specific summary. Also, we wish to add an interactive program for answering questions about the results. For example, an investigator should be able to find out easily how many processes involve cleavage of a specific bond and how strong their resulting MS peaks are.

The INTSUM program is now used routinely by mass spectroscopists at Stanford engaged in investigations of the mass spectrometric fragmentation of various classes of organic compounds, primarily steroids. A manuscript is now in preparation (Ref. 54) describing the fragmentation of progesterone and related compounds. The program was used extensively in this work. We are now beginning a detailed examination of the fragmentation of steroids related to the androstane skeleton, particularly the biologically important testosterones. We propose to continue to use the INTSUM program in its present form and as it is improved in support of these studies.

The generator of rules that we now have, RULEGEN, does a credible job of explaining the regularities summarized by INTSUM. It has found, for example, the well-known alpha-cleavage fragmentation process and beta cleavage followed by rearrangement in the low resolution data for fifteen aliphatic amines. The program will be extended in two important ways to increase its utility: (i) the program needs to be able to work with an increased number of descriptive predicates in the generation of rules, and (ii) it needs to be given a more flexible representation of complex fragmentation mechanisms so that it can more easily find rules involving more than two bonds.

We will continue working with low resolution MS data of the 150-200 monofunctional aliphatic compounds studied previously in the context of the performance program. These compounds are well-understood and thus provide a good test of the program's effectiveness. In order to insure generality in the theory formation programs, we will also test the system against the high resolution mass spectra of the 68 estrogenic steroids. Since they are also well-understood, these compounds will show how well the program can deal with complex ring systems, multifunctional compounds, cleavages involving more than two bonds, and high resolution data.

The existing programs are in good working order - within definite limits - so we expect to apply them to new sets of data from the MS laboratory as interest arises. For example, as the high and low resolution MS from marine sterols are collected we expect to use INTSUM and RULEGEN (at least) to assist in the interpretation and generalization of these data. Since these problems will advance the state of knowledge of MS, it is not correct to look on them as test problems. However, in the past the programs developed most rapidly when they were applied to unsolved problems of interest to our colleagues in the chemistry department.

For development of the interactive programs, we will rely heavily on the criteria of acceptability by Stanford users. The programs themselves will be written in INTERLISP on the SUMEX computer. Initially, we will provide interactive access to the control parameters of the programs in order to allow users to tailor their

runs to their immediate interests. Later we hope to expand these to allow interrogation of the programs with respect to both contents of the results and the program's reasoning steps.

D. Applications to Biomedical Problems

We can immediately offer to the user community the Planner, for analysis of HR/MS in terms of molecular structure. The program is insensitive to the source of the MS data, and we foresee significant use of the program for analysis of spectra of mixtures without prior separation and spectra from the GC/HRMS facility without additional programming effort. Examples of applications areas are summarized below.

We wish to exploit our existing capabilities of the analysis of biological mixtures without prior separation (ref. 33). This approach will prove particularly useful in studies of mixtures which are difficult to separate and analyze by GC. Phytoecdysones related to ecdysone, an insect molting hormone, present such a problem. GC of these compounds is very difficult, although high-pressure liquid chromatography has recently been used to carry out separations. This class of compounds represents an interesting and valuable test case for our combined MS and computer techniques, particularly the specification and subsequent acquisition of metastable defocussing data for precise linking of parent and fragment ions in the spectrum of a complex mixture (refs. 28, 33). Model compounds, mixtures and current structure elucidation problems are available (Nakanishi, Columbia; Takemoto, Tohoku University, Sendai, Japan). Although most users cannot be completely specific as to the nature of their future structure elucidation problems, we feel that several of these problems can be handled by such an approach.

As the structure generator and its extensions are developed further, we foresee continuing use of an interactive version applied to specific problems of the user community. As an example, the work in collaboration with the GRC project will involve studies of several classes of compounds extracted from human body fluids (e.g., aromatic and aliphatic acids, various classes of bases, amino acids and carbohydrates) which contain representatives varying by substitutions about a small number of molecular skeletons. The generator can define all isomers which must be considered as possible solutions.

For those problems which are amenable to attack by library search procedures, e.g., screening of GC/LRMS runs of marine sterols to weed out known compounds, we propose to use these procedures and to investigate extensions to them. using a procedure related to that described by McLafferty (K-S. Kwok, et al., J Amer. Chem. Soc., 95, 4185 (1973), we seek to determine from modified library search techniques the known structures which yield similiar spectra. Utilizing the DENDRAL structural manipulation routines, we would then seek to determine those related structures (whose spectra are not in the library) which are possible solutions. A library, including Wiswesser Line Notation names, exists (F. W. McLafferty, private communication) and would be of some utility in this work.

The MS facility in conjunction with our programs will be used in studies of the following nature:

1) Prof. Djerassi - we plan use of the MS facilities and computer programs in ongoing research connected with existing NIH-supported studies on steroids and marine sterols and continued collaboration with Prof. Adlercreutz on estrogen mixtures isolated from body fluids. Further collaboration with Prof. Adlercreutz will be on structural studies of new estrogen metabolites whose presence in mixtures has been inferred through our previous collaborative efforts.

The work on marine sterols presently utilizes GC/LRMS and frequently laborious separation procedures to isolate individual fractions for HRMS analysis. GC/HRMS will be a significant assistance in this effort. We plan MS studies of known marine sterols (utilizing INTSUM) to derive fragmentation rules, which then will be used in the Planner to aid structure elucidation of new compounds.

We also plan further work on extensions of MS theory in the steroid field, initially focussed on additional biomedically important classes of steroids related to the pregnane (progesterones) and androstane (testosterones) skeletons. This work is currently being carried out by Dr. Smith in collaboration with two visiting senior scientists (Dr. Roy Gritter, Dr. Geoff Dromey) currently on sabbatical leave fellowships.

2) Chemistry Department Collaborators - as indicated by the responses summarized in the letters of interest (Appendix A), there is significant interest in use of the MS facility by other NIH-supported members of the chemistry department. All those listed are familiar with the technique of MS as applied to structure elucidation problems. Most have used MS frequently, particularly Prof. Van Tamelen in his studies of the cyclization of squalene and related studies in the terpenoid and steroid field. The interests of these collaborators are generally in HRMS and GC/HRMS, with occasional use of other capabilities of the system. The types of compounds studied by this group and an indication of the amount of use expected are summarized in the letters of interest.

3) Genetics Research Center (GRC): (Profs. J. Lederberg, H. Cann; Dr. A. Duffield)

The body fluids analyzed by GC/LRMS to date include urine, blood, amniotic fluid and cerebrospinal fluid. Each body fluid is fractionated into the following compound classes:

a) organic acids and neutral compounds
b) amino acids
c) carbohydrates

which after appropriate derivatization are analyzed by GC/LRMS/computer system. A library of known LRMS will serve as the primary means of identifying metabolites from their experimentally recorded LRMS.

In those instances where the LRMS is insufficient for metabolite identification GC/HRMS data will be necessary to determine the composition of all ions in its mass spectrum. These data will greatly enhance the prospects of identifying the metabolite in question.

It is known (on past performance) that if a compound is present in body fluids at the level of 1 microgram per GC peak then good quality HR/MS will be recorded (ion amplitude dynamic range of 1:100, mass accuracy of +-5ppm) using the Varian MAT 711 mass spectrometer. If the GC peak of interest contains insufficient material for a HRMS scan then preparative GC could be used to concentrate that portion of the chromatograph effluent prior to GC/HRMS.

Prior to the demise of the ACME computer system (July 31, 1973) we developed a GC/HRMS system and applied it to the analysis of extracts from body fluids. The following example represents results obtained with this system during its development. The example used was a routine analysis and was run to determine the capability of the overall system during its development and not as an unknown sample of extreme interest.

The total ion plot recorded during the lifetime of the GC/HRMS analysis of an amniotic fluid is reproduced as Figure 1. A complete high resolution scan was recorded on each of the peaks shown in Figure 1. Filing time of the time-shared ACME computer system did not allow the system to operate in a repetitive scan mode. For the sake of brevity only the GC/HRMS scan (# 1594, Figure 2) corresponding to glutamic acid N-TFA O-n-Butyl ester derivative is produced. (The corresponding GC/LRMS scan is Figure 3). The scan time per decade of mass was 10.5 seconds, the resolution 6,500 and the matching tolerance for the assignment of empirical composition set to 4 mmu. The results show that the system was capable of accurate mass measurement with a dynamic range in ion amplitude of about 33:1 in this instance.

The cessation of computer support for the GC/HRMS system did not allow a HRMS analysis to be made which was crucial to the identification of a metabolite present in a body fluid. Since that time however, several instances have arisen where GC/HRMS data would have been collected in an effort to identify metabolites not previously seen.

The expected sample throughput in the GRC project with existing personnel is expected to approach 5 to 7 body fluids per week (15-21 GC/LRMS fractions to be run in the Genetics Department per week). On average GC/HRMS would be required on 1 - 2 samples per week.

The research interests of the Medical School collaborators relative to the proposed MS resource are summarized in the letters of interest (Appendix A). The MS services required by this community will include GC/LRMS (Forrest, Sera, Kalman for drug and drug metabolite identification, Rabinowitz and Wilkinson for prostaglandin identification, Robin for identification of oxidized/reduced redox pairs, Hollister for Marihuana metabolites, Barchas, neurotransmitters, Fair, polyamines and the prostatic antibacterial factor in urine); GC/HRMS (Trudell, drug metabolite identification, Kvenvolden, structure of amino acids and related compounds plus samples as required from interests described under GC/LRMS). In those instances where the biological extract contains insufficient material for a GC/HRMS scan preparative GC, using existing instrumentation within the chemistry department, can be

used to concentrate the material prior to the GC/HRMS analysis. If the material of interest is obtained relatively pure by this technique then HRMS analysis using direct sample insertion into the ion source would be utilized.

As mentioned above, several of the computer programs have immediate utility for assisting with structure elucidation problems. For example, the Structure Generator program can answer structural isomerism questions independently of mass spectrometry, (e.g. , to provide lists of isomers in conjunction with isomer interconversion problems such as carbonium ion rearrangements). Because the program will be able to generate complete lists of isomers with (or without) some specified structural features, a researcher can have confidence that no possibilities have been overlooked. Some interest in the structure generator has been expressed by representatives of the pharmaceutical industry. The generator could be used to suggest complete sets of structural alternatives for possible synthesis, once a physiologically active congener has been identified.

In more general terms, the structure generator can be richly suggestive of new, unexplored areas of synthetic organic chemistry. for example, the generator has been used by a graduate student in chemistry, Mr. Jan Simek, to identify the space of possible Diels-Alder condensation products consisting of six atoms of any combination of carbon, nitrogen, oxygen, and sulfur in a six-membered ring with one double bond. A literature search through the Ring Index revealed that many of the ring systems have never been reported.

IV. SIGNIFICANCE OF PROPOSED RESEARCH

Structure elucidation is an important and difficult problem for biomedical scientists. Many of them lack the detailed chemical background necessary to be efficient in this endeavor. Generally speaking, they also lack the frequently complex and expensive equipment (e.g., high resolution mass spectrometers) to provide spectroscopic data to assist them in solving problems of molecular structure. We plan to provide the chemical and analytical expertise to facilitate the solution of their structural problems. This research aims at providing more powerful techniques for determining molecular structures than are now routinely available. In particular, we have proposed (a) providing extended MS services as a means of collecting powerful analytic data for scientists; (b) developing (and extending) sophisticated computer programs to assist with the interpretation of the data from mass spectrometry and elsewhere, (c) developing (and extending) novel computer programs to assist with formulation of the rules of interpretation, and (d) applying these state of the art techniques to problems of biomedical relevance. Our research group is thus dedicated to a broad-based attack on the applications of structure elucidation to biological and biomedical problems.

The proposed research not only holds promise for significant long-term advances, it can have immediate benefits as well. Many members of the biomedical community at Stanford have called upon the MS laboratory for assistance in the past and will continue to do so in the future. The proposed resource will provide the conduit for a substantial increase in the utilization of MS within the Stanford biomedical community. The ability of the proposed resource to interpret the experimental data it generates (enhanced by the close proximity of the resource and biomedical community) should result in a successful program of interdisciplinary research.

HRMS is an important source of data for these problems, and GC/HRMS is still more important. Previous investment by the NIH in the Varian MAT-711 HRMS system at Stanford can be utilized now and built upon for the future. Continued operation of the GC/MS system will give the Stanford community access to state-of-the-art spectroscopic techniques and to professional mass spectroscopists who can help with ongoing problems.

The computer programs themselves constitute a unique resource for assisting with the structure determination. The previous NIH grant supported development of the programs. In part, we are requesting funds to exploit these programs.

One of the most significant aspects of this work is its interdisciplinary view of solving molecular structure problems by intelligently directed search of the space of chemical graph structures. As a result of posing the structure determination problem in this framework, we have been able to further the knowledge about structure elucidation in at least three ways. First, some of the knowledge used by analytical chemists has been made more precise for use in a computer program. Second, codifying such knowledge for the computer has led to the discovery

of new research areas to extend our existing knowledge of MS. Several publications listed in the bibliography (Refs. 42 and following) are reports of exactly this kind of research. Finally, the computer's systematic search through the space of possible structures gives the practicing scientist the confidence that no structures were merely overlooked. The efficiency of the program depends on the exclusions of many whole classes of compounds, but the computer will have rejected those classes using precise, explicitly stated criteria.

Our recent work on finding MS interpretation rules (theory formation) can provide additional unique capabilities for assisting with the problem solving. We wish to continue this research because it offers hope for a solution to the problem of furnishing real-world knowledge to computer programs -- in particular to the computer programs that assist with structure elucidation. This is a pressing problem in current AI research. High performance programs, of which DENDRAL is most often cited, derive their power from large stores of knowledge. Yet there are no routine methods for infusing such systems with knowledge of the task domain. We believe our research in theory formation holds a key to the solution of this problem.

V. FACILITIES & EQUIPMENT

The Stanford Mass Spectrometry Laboratory will provide MS services on the Varian MAT-711 mass spectrometer coupled with a Hewlett-Packard gas chromatograph (Model 7610A). As service instruments for more routine mass spectral analyses, the laboratory has a MS-9 and CH-4 mass spectrometers.

Data reduction is currently provided on Stanford's IBM 370/158 computer in conjunction with a front-end PDP-11/20 data acquisition computer. (The PDP-11/20 presently has only the capability for buffering peak profile data between the mass spectrometer and the IBM 370/158 computer at the Stanford Computer Center.) An alternative to buying time on the 370/158 is proposed and discussed in the budget justification.

The AI programs will be run on the NIH-sponsored AIM-SUMEX computer facility (a PDP-10 computer with the TENEX operating system, 192K words of memory, and adequate peripherals for our purposes). Running these programs on SUMEX will incur no charge.

VI. BIBLIOGRAPHY

A. 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

(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, 91:11 (May 21, 1969).

(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", in R. Banerji & M.D. Mesarovic (eds.) Theoretical Approaches to Non-Numerical Problem Solving, New York: Springer-Verlag, 1970. (Also, 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, 91:26 (December 17, 1969).

(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 Chemica 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. Buchanan, E.A. Feigenbaum, J. Lederberg, G. Schroll, and G.L. Sutherland, "The Application of Artificial Intelligence in the Interpretation of Low- Resolution Mass Spectra", Advances in Mass Spectrometry, 5, 314.

(25) B.G. Buchanan and J. Lederberg, "The Heuristic DENDRAL Program for Explaining Empirical Data". In proceedings of the IFIP Congress 71, Ljubljana, Yugoslavia (1971). (Also Stanford Artificial Intelligence Project Memo No. 141.)

(26) B.G. Buchanan, E.A. Feigenbaum, and J. Lederberg, "A Heuristic Programming Study of Theory Formation in Science." In proceedings of the Second International Joint Conference on Artificial Intelligence, Imperial College, London (September, 1971). (Also Stanford Artificial Intelligence Project Memo No. 145.)

(27) Buchanan, B. G., Duffield, A.M., Robertson, A.V., "An Application of Artificial Intelligence to the Interpretation of Mass Spectra", Mass Spectrometry Techniques and Appliances, Edited by G. W. A. Milne, John Wiley & Sons, Inc., 1971, p. 121-77.

(28) D.H. Smith, B.G. Buchanan, R.S. Engelmore, A.M. Duffield, A. Yeo, E.A. Feigenbaum, J. Lederberg, and C. Djerassi, "Applications of Artificial Intelligence for Chemical Inference VIII. An approach to the Computer Interpretation of the High Resolution Mass Spectra of Complex Molecules. Structure Elucidation of Estrogenic Steroids", Journal of the American Chemical Society, 94, 5962-5971 (1972).

(29) B.G. Buchanan, E.A. Feigenbaum, and N.S. Sridharan, "Heuristic Theory Formation: Data Interpretation and Rule Formation". In Machine Intelligence 7, Edinburgh University Press (1972).

(30) Lederberg, J., "Rapid Calculation of Molecular Formulas from Mass Values". Jnl. of Chemical Education, 49, 613 (1972).

(31) Brown, H., Masinter L., Hjelmeland, L., "Constructive Graph Labeling Using Double Cosets". Discrete Mathematics (in press). (Also Computer Science Memo 318, 1972).

(32) B. G. Buchanan, Review of Hubert Dreyfus' "What Computers Can't Do: A Critique of Artificial Reason", Computing Reviews (January, 1973). (Also Stanford Artificial Intelligence Project Memo No.

(33) D. H. Smith, B. G. Buchanan, R. S. Engelmore, H. Aldercreutz and C. Djerassi, "Applications of Artificial Intelligence for Chemical Inference IX. Analysis of Mixtures Without Prior Separation as Illustrated for Estrogens". Journal of the American Chemical Society 95, 6078 (1973).

(34) D. H. Smith, B. G. Buchanan, W. C. White, E. A. Feigenbaum, C. Djerassi and J. Lederberg, "Applications of Artificial Intelligence for Chemical Inference X. Intsum. A Data Interpretation Program as Applied to the Collected Mass Spectra of Estrogenic Steroids". Tetrahedron, 29, 3117 (1973).

(35) B. G. Buchanan and N. S. Sridharan, "Rule Formation on Non-Homogeneous Classes of Objects". In proceedings of the Third International Joint Conference on Artificial Intelligence (Stanford, California, August, 1973). (Also Stanford Artificial Intelligence Project Memo No. 215.)

(36) D. Michie and B.G. Buchanan, "Current Status of the Heuristic DENDRAL Program for Applying Artificial Intelligence to the Interpretation of Mass Spectra". August, 1973.

(37) H. Brown and L. Masinter, "An Algorithm for the Construction of the Graphs of Organic Molecules", Discrete Mathematics (in press). Also Stanford Computer Science Department Memo STAN-CS-73-361, May, 1973)

(38) D.H. Smith, L.M. Masinter and N.S. Sridharan, "Heuristic DENDRAL: Analysis of Molecular Structure," Proceedings of the NATO/CNA Advanced Study Institute on Computer Representation and Manipulation of Chemical Information, in press.

(39) R. Carhart and C. Djerassi, "Applications of Artificial Intelligence for Chemical Inference XI: The Analysis of C13 NMR Data for Structure Elucidation of Acyclic Amines", J. Chem. Soc. (Perkin II), 1753 (1973).

(40) L. Masinter, N. Sridharan, and D.H. Smith, "Applications of Artificial Intelligence for Chemical Inference XII: Exhaustive Generation of Cyclic and Acyclic Isomers.", submitted to Journal of the American Chemical Society.

(41) L. Masinter, N. Sridharan, R. Carhart and D.H. Smith, "Applications of Artificial Intelligence for Chemical Inference XIII: An Algorithm for Labelling Chemical Graphs", submitted to Journal of the American Chemical Society.

(42) The Determination of Phenylalanine in Serum by Mass Fragmentography. Clinical Biochem., 6 (1973). By W.E. Pereira, V.A. Bacon, Y. Hoyano, R. Summons and A.M. Duffield.

(43) The Simultaneous Quantitation of Ten Amino Acids in Soil Extracts by Mass Fragmentography. Anal. Biochem., 55, 236 (1973). By W.E. Pereira, Y. Hoyano, W.E. Reynolds, R.E. Summons and A.M. Duffield.

(44) An Analysis of Twelve Amino Acids in Biological Fluids by Mass Fragmentography. Anal. Chem., in press. By R.E. Summons, W.E.

Pereira, W.E. Reynolds, T.C. Rindfleisch and A.M. Duffield.

(45) The Quantitation of B-Aminoisobutyric Acid in Urine by Mass Fragmentography. Clin. Chim. Acta, in press. By W.E. Pereira, R.E. Summons, W.E. Reynolds, T.C. Rindfleisch and A.M. Duffield.

(46) The Determination of Ethanol in Blood and Urine by Mass Fragmentography. Clin. Chim. Acta, in press. By W.E. Pereira, R.E. Summons, T.C. Rindfleisch and A.M. Duffield.

Publications Describing DENDRAL-Related Research But Not Funded By This Grant

(47) An Automated Gas Chromatographic Analysis of Phenylalanine in Serum. Clinical Biochem., 5, 166 (1972). By E. Steed, W. Pereira, B. Halpern, M. D. Solomon and A.M. Duffield.

(48) Pyrrolizidine Alkaloids. XIX. Structure of the Alkaloid Erucifoline. Coll. Czech. Chem. Commun., 37, 4112 (1972). By P. Sedmera, A. Klasek, A.M. Duffield and F. Santavy.

(49) Chlorination Studies I. The Reaction of Aqueous Hypochlorous Acid with Cytosine. Biochem. Biophys. Res. Commun., 48, 880 (1972). By W. Patton, V. Bacon, A.M. Duffield, R. Halpern, Y. Hoyano, W. Pereira and J. Lederberg.

(50) A Study of the Electron Impact Fragmentation of Promazine Sulphoxide and Promazine using Specifically Deuterated Analogues. Austral. J. Chem., 26, 325 (1973). By M.D. Solomon, R. Summons, W. Pereira and A.M. Duffield.

(51) Spectrometrie de Masse VIII. Elimination d'can Induite par Impact Electronique dans Le Tetrahydro-1,2,3,4-Napthtal-ene-diol-1,2. Org. Mass Spectre., 7, 357 (1973). By P. Perros, J.P. Morizur, J. Kossanyi and A.M. Duffield.

(52) Chlorination Studies II. The Reaction of Aqueous Hypochlorous Acid with a-Amino Acids and Dipeptides. Biochim. et Biophys. Acta, 313, 170 (1973). By W.E. Pereira, Y. Hoyano, R. Summons, V.A. Bacon and A.M. Duffield.

(53) Spectrometrie de Masse. IX. Fragmentations Induites par Impact Electronique de Glycols- En Serie Tetraline. Bull. Chim. Soc. France, 2105 (1973). By P. Perros, J.P. Morizur, J. Kossanyi and A.M. Duffield.

(54) The Use of Mass Spectrometry for the Identification of Metabolites of Phenothiazines. Proceedings of the Third International Symposium on Phenothiazines, Raven Press, New York, 1973. By A.M. Duffield.

(55) Chlorination Studies IV. The Reaction of Aqueous Hypochlorous Acid with Pyrimidine and Purine Bases. Biochem. Biophys. Res. Commun., 53, 1195 (1973). By Y. Hoyano, V. Bacon, R.E. Summons, W.E. Pereira, B. Halpern and A.M. Duffield.

(56) Mass Spectrometry in Structural and Stereochemical Problems. CCXXXVII. Electron Impact Induced Hydrogen Losses and Migrations in Some Aromatic Amides. Org. Mass Spectry., in press. By A.M. Duffield, G. deMartino and C. Djerassi.

(57) Stable Isotope Mass Fragmentography: Quantitation and Hydrogen-Deuterium Exchange Studies of Eight Murchison Meteorite Amino Acids. Geochem. et Cosmochim. Acta, submitted for publication. By W.E. Pereira, R.E. Summons, T.C. Rindfleisch,

A.M. Duffield, B. Zeitman and J.G. Lawless. (58) Mass Spectrometry in Structural and Stereochemical Problems CLXXXIII. A Study of the Electron Impact Induced Fragmentation of Aliphatic Aldehydes. J. Amer. Chem. Soc., 91, 6814 (1969). By R.J. Liedtke and C. Djerassi.

(59) Mass Spectrometry in Structural and Stereochemical Problems - CXCVII. Electron-Impact Induced Functional Group Interaction in 4-Benzyloxycyclohexyl Trimethylsilyl Ether. Org. Mass Spectrom. 4, 257 (1970). By Paul D. Woodgate, Robin T. Gray and Carl Djerassi.

(60) Mass Spectrometry in Structural and Stereochemical Problems - CXCVIII. A study of the Fragmentation Processes of Some a,B-Unsaturated Aliphatic Ketones. Org. Mass Spectrom., 4, 273 (1970). By Younus M. Sheikh, A.M. Duffield and Carl Djerassi.

(61) Mass Spectrometry in Structural and Stereochemical Problems CCII. Interaction of Remote Functional Groups in Acyclic Systems upon Electron Impact. J. Org. Chem., 36, 1796 (1971). By M. Sheehan, R.J. Spangler, M. Ikeda and C. Djerassi.

(62) Mass Spectrometry in Structural and Stereochemical Problems CCVII. Fragmentation of Unsaturated Ethers. Org. Mass Spectrom., 5, 895 (1971). By J. P. Morizur and C. Djerassi.

(63) Mass Spectrometry in Structural and Stereochemical Problems CCVIII. The Effect of Double Bonds Upon the McLafferty Rearrangement of Carbonyl Compounds. J. Amer. Chem. Soc., 94, 473 (1972). By J.R. Dias, Y.M. Sheikh and C. Djerassi.

(64) Mass Spectrometry in Structural and Stereochemical Problems CCXV. Behavior of Phenyl-Substituted a,B-Unsaturated Ketones Upon Electron Impact. Promotion of Hydrogen Rearrangement Processes. J. Org. Chem., 37, 776 (1972). By R.J. Liedtke, A.F. Gerrard, J. Diekman and C. Djerassi.

(65) Mass Spectrometry in Structural and Stereochemical Problems CCXXII. Delineation of Competing Fragmentation Pathways of Complex Molecules from a Study of Metastable Ion Transitions of Deuterated Derivatives. Org. Mass Spectrom., 7, 367 (1973). By D.H. Smith, A.M. Duffield and C. Djerassi.

(66) The Carbon-13 Magnetic Resonance Spectra of Acyclic Aliphatic Amines. J. Amer. Chem. Soc., 95, 3710 (1973). By H. Eggert and C. Djerassi.

(67) The Carbon-13 Nuclear Magnetic Resonance Spectra of Keto Steroids. J. Org. Chem., 38, 3788 (1973). By H. Eggert and C. Djerassi.

(68) Mass Spectrometry in Structural and Stereochemical Problems CCXXXVIII. The Effect of Heteroatoms Upon the Mass Spectrometric Fragmentation of Cyclohexanones. J. Org. Chem., in press. By J.H. Block, D.H. Smith, and C. Djerassi.

(69) Mass Spectrometry in Structural and Stereochemical Problems CCXLII. Applications of DADI, a Technique for Study of Metastable Ions, to Mixture Analysis. J. Amer. Chem. Soc., submitted for publication. By D.H. Smith, C. Djerassi, K.H. Maurer, and U.

Rapp.

(70) Mass Spectrometry in Structural and Stereochemical Problems. The Fragmentation of Progesterone and Alkyl-Substituted Progesterones, in preparation, by S. Hammerum and C. Djerassi.

(71) Applications of Molecular Orbital Theory to the Interpretation of Mass Spectra: Prediction of Primary Fragmentation Sites in Organic Molecules," Org. Mass Spectrom., 7, 1241 (1973), by G. Loew, M. Chadwick, and D.H. Smith.

PROGRESS REPORT

TEXT OF 1973 ANNUAL REPORT FOR RESEARCH
PROJECT: RESOURCE-RELATED RESEARCH -- COMPUTERS
AND CHEMISTRY

Progress Report

Part A. APPLICATIONS OF ARTIFICIAL INTELLIGENCE TO MASS SPECTROMETRY

OBJECTIVES:

Research activities carried out under Part A of this project have been directed toward extending the reasoning power of Heuristic DENDRAL. Heuristic DENDRAL reresents a paradigm for attacking problems in one of the major areas of importance to any scientific discipline dealing with molecules, the area of structure elucidation. We have focused our attention on the use of heuristic programming techniques for analysis of mass spectra and ancillary analytical data which can be obtained utilizing a mass spectrometer. It is convenient to discuss objectives, progress and plans by examining three broad areas of activity in research connected with Part A. We wish to note that these areas conform to our overall strategy of PLAN-GENERATE-TEST. We have shown, earlier, how powerful this strategy is when applied to the task of structure elucidation utilizing mass spectral data. The areas and their objectives are the following:

(I) PLANNER:
- (a) Extend the programs used for structure elucidation to structural analysis of complex molecules.
- (b) Assess the capabilities and limitations of the PLANNER.
- (c) Generalize the programming techniques to reduce compound class dependence.
- (d) Explore the utility of ancillary data available from the mass spectrometer.

(II) STRUCTURE GENERATOR:
- (a) Complete the exhaustive, irredundant generator of molecular structures.
- (b) Develop efficient constraints on the generator to exploit its potential utility.
- (c) Exploit the concepts developed for the structure generator in solving various structure-problems (related to m.s. and others) and isomer-problems.

(III) PREDICTOR
- (a) Extend the Predictor to still more complex molecular structures.
- (b) Explore the design of experimental strategies, utilizing Predictor functions, to differentiate among candidate solutions.

We point out that the PLAN-GENERATE-TEST strategy, although applied to structure elucidation, has potential utility as a strategy for solving other chemical problems. Similarly, although we utilize mass spectral data almost exclusively, the same heuristic programming techniques allow facile extension to analysis of data from other types of analytical instrumentation. These were not objectives of the original research proposal but seem logical extensions for future work. We have illustrated the potential of these techniques for analysis of 13C NMR data (Carhart and Djerassi, 1973). This is discussed briefly under the PLANS section, below.

PROGRESS:

(I) PLANNER

The function of the Planner is to analyze mass spectral data acquired on a compound. The Planner attempts to derive structural information from these data using the rules of behavior of compounds in the mass spectrometer.

Objective (a): Extend Programs.

The Planner is presently embodied in a program which also contains a set of functions to assemble this structural information into complete molecules (a primitive Structure Generator) and to test these molecular structures with other, not necessarily mass spectral, rules (a primitive Predictor). This performance program was written in this way to provide a useful tool for chemical studies while more general versions of the Structure Generator and Predictor were being developed. This program and its performance have been described in some detail in a publication and in previous progress reports. A manuscript (Smith, et.al., 1973) has now appeared describing the application of this program to the analysis of mixtures of compounds without prior separation.

Objective (b): Assess Capabilities.

We have extended the capabilities of the Planner so that we can analyze both low and high resolution mass spectral data. A low resolution mass spectrum is regarded by the program as a pseudo high resolution spectrum wherein possible elemental compositions of each peak are limited only by the inferred molecular formula of the compound. This results in more ambiguity with a commensurate increase in number of candidate solutions as would be expected considering the lower specificity of low resolution data as compared to high resolution data.

We have extended our capabilities for molecular ion determination utilizing a heuristic search technique through the space of plausible molecular ions. This technique has had significant success even when dealing with the low resolution mass spectra of compounds which display no molecular ion, for example the class of derivatized amino acids (trifluoracetyl, n-butyl esters) important to studies carried out under Part B, below.

We have segmented the performance program to decrease the amount of memory required for its operation. This should increase the chances for other groups to make use of the program.

The limitations of the present performance program are primarily the requirement that some information about the class of compounds be known, and that, for each class, relatively detailed rules about the mass spectral fragmentation of this class be available. The former limitation results primarily from the nature of the program in that a complete structure generator is not incorporated. The primitive structure generator available to the program can only place substituents about an assumed skeleton. This limitation will be alleviated when a structure generator with GOODLIST and BADLIST constraints is available (see Structure Generator, below). The latter limitation is more fundamental, but is characteristic of every spectroscopic technique to one degree or another. It must be assumed that analysis of a mass spectrum, alone, may not lead to sufficiently unambiguous information about the structure of the compound yielding the spectrum. It is for this reason that extensions of the programming techniques to encompass data from other spectroscopic techniques are attractive.

Objective (c): Generalize Techniques.

We have carried out several successful experiments to ensure that the performance program, used originally for analysis of estrogenic steroids, retains only procedures which are compound-class independent. By supplying fragmentation rules for other classes of compounds, we have successfully carried out structure elucidation of molecules in several diverse classes including other steroidal hormones and related compounds (progesterones, testosterones, androsterones), steroidal sapogenins and derivatized amino acids.

Objective (d): Explore Utility

Previous progress reports have summarized in some detail the ways in which data from ancillary techniques in mass spectrometry (metastable ion and low ionizing voltage data, labile hydrogen exchange) can be used by the program. The utility of metastable ions for aid in stucture elucidation continues as an active area of interest. Experience with the program has inspired studies on metastable ions, first, to help delineate the course of fragmentation of molecules with the purpose of extending and refining fragmentation rules used by the program (Smith, Duffield and Djerassi, 1973). Experience with the increased specificity of structural information with concomitant reduction in analysis time when metastable ion information is available (Smith, et.al., 1973) has led to a study of a new technique for detection and analysis of metastable ions (Direct Analysis of Daughter Ions, or DADI) and has illustrated the utility of this technique in mixture analysis (Smith, Djerassi, Maurer and Rapp, 1973).

Experience with the PLANNER has led to several research activities related to, but not supported by, this grant. Our studies of estrogen mixtures isolated from pregnancy urine have suggested new compounds likely to be important in the human metabolism of estrogens. Some of these compounds are hitherto unreported structures and a synthesis program is underway in Professor Djerassi's laboratory to produce some of these compounds. The Planner will be used as one method of comparison of the synthesized, authentic standards with those isolated from pregnancy urine.

Work is also being carried out to explore the fragmentation of model systems possessing two heteroatoms in close proximity. It is clear from the first of these studies (Block, Smith, and Djerassi, 1973) that the fragmentation of these difunctional systems does not reflect that of monofunctional analogs. More groundwork is required in this area to obtain better fragmentation rules for these systems.

II. STRUCTURE GENERATOR

Objective (a): Complete the Generator

The last progress report discussed the completion of both the basic structure generator algorithm and program, which provide the capability for exhaustive generation of graph isomers of a given empirical formula, with prospective avoidance of duplicate structures. Since the time of the submission of that report, manuscripts describing the structure generator, directed specifically to an audience of chemists, have been submitted (Masinter, Sridharan, Lederberg, and Smith, 1973; Masinter and Sridharan, 1973). Some effort over the past year has been devoted to

verification of the completeness and irredundancy of the method. We have extended existing combinatorial counting algorithms to check that the numbers of isomers generated are correct. We have used an interactive version of the generator to verify that variations (allowed by the algorithm) of the mechanism of generation yield the same set of isomers. In this way we are now increasingly confident that the program's performance accurately reflects the mathematically proven algorithm on which it is based.

The Structure Generator has been briefly described, and placed in its context within Heuristic DENDRAL, in an invited paper presented at a NATO/CRS sponsored conference on Computer Representation and Manipulation of Chemical Information, held in Amsterdam in June, 1973 (Smith, Masinter and Sridharan, 1973).

We have also begun to develop techniques to expand the scope of the generator. One example, which has been completed, is adding extensions to the CATALOG. The CATALOG contains the set of vertex-graphs from which structures are assembled. The original CATALOG was not sufficient to generate all isomers of some potentially interesting compositions, e.g., those involving graphs possessing nodes of degree >3. We now have a program which constructs complete sets of vertex-graphs containing nodes of degree >3 from the set of trivalent graphs in the original CATALOG. We have thus extended the capabilities of the generator. Other such extensions are discussed in the PLANS section, below.

Objective (b): Develop Constraints

It is absolutely essential that we provide the mechanism for constraining the Structure Generator: without constraints it is merely a legal move generator, as in a chess-playing program. For structure elucidation problems, the Planner can determine many features of the molecular structure from various types of experimental data such as presence of functional groups, and the numbers of double bonds and rings. Partial information of this sort can be used to constrain the Structure Generator to the space of plausible candidate structures. From a graph-theoretic point of view, however, constraining the graph generating algorithm is a difficult unsolved problem.

We are presently formulating several types of constraints to apply to the Structure Generator. Some types of constraint await the development of new mathematical tools (see PLANS), while others can be immediately implemented with relatively minor alterations to the algorithm. The class of constraints presently receiving attention deals with types of unsaturation (rings or double bonds) desired in the final structures. Related to this constraint is the constraint of number of quaternary carbons present. The former information (number and nature of multiple bonds) is readily available from several spectroscopic techniques, while the latter may be obtained from 13C NMR. The implementation of this class of constraints will be used as the model for future implementation of a GOODLIST (structural features known to be present) and a BADLIST (structural features known to be absent).

It is possible that some types of constraints may not be easily implemented within the algorithm. Thus, retrospective tests of isomers may be required to search for desired or unwanted features. We have developed some new approaches to graph matching which seem to be significantly more efficient than previous methods. Should prospective implementation of a constraint prove difficult, we will

have at our disposal some powerful graph matching tools to exercise the constraint.

Objective (c): Exploit the Generator for Structure Elucidation

We have demonstrated the utility of some subsystems of the structure generator, e.g., the LABELLER, by exploring some problems of isomerism noted in the chemical literature. We have corrected the member and presented the identities of isomers formed by different substitutions of alkyl chains about a porphyrin nucleus. We are presently exploring some problems of isomerism of carbocyclic ring systems, specifically C10H10 and (CH)10 and C10H2n-4 tricyclic ring systems, n = 8 - 12, related to the mechanistics of isomeric interconversion.

We have the complete list of all topologically possible 1176 6-membered Diels-Alder ring systems, using any combination of C,N,O and S. This list was generated using the PARTITIONER and an extended version of the LABELLER. These are all the 6-membered ring systems that can be embedded in structures resulting from the well-known Diels-Alder reaction. Of the 1176 possible ring systems, approximately 80% are unreported in the Ring Index. Many of these are chemically unstable - underscoring the need for a BADLIST implementation for the Cyclic Structure Generator. However, many of these unreported ring systems are certainly chemically plausible. Awareness of such gaps in relatively simple synthetic categories might lead to discovery of new categories of compounds with important biological effects.

(III) PREDICTOR

The function of the Predictor in the PLAN-GENERATE-TEST strategy is to perform a detailed evaluation of candidate solutions (structures) to a structure elucidation problem. It may use a more detailed model of spectroscopic behavior than that embodied in a Planner to attempt to differentiate among possible solutions.

Objective (a): Extend the Program

We have extended and generalized the Predictor used previously for saturated, aliphatic, monofunctional compounds. Given a list of structures and rules of fragmentation processes, it will predict a mass spectrum for each structure. Prediction of relative ion abundances is crude, but previous work has shown that even crude measures of ion abundance are usually satisfactory. The predicted spectrum can be matched then with the observed and candidates ranked according to the quality of the match. The program works with structures and rules of any complexity. An interesting philosophical question is how much knowledge should be brought to bear on interpretation of the data at the Planning vs. Predicting stages of analysis. It is our feeling that if more can be accomplished during Planning to constrain the Structure Generator, the analysis will be more efficient. On the other hand, some knowledge can be utilized only if a complete structure is specified, so that its use is restricted to a predictive role. Moreover, Predictor Functions have a greater utility, as indicated in the subsequent section.

Objective (b): Differentiate Structures

The Predictor has a more obvious application in the design of

experimental strategies to differentiate among candidate structures. Rules of spectroscopic behavior utilized during Planning demand the presence of some data to evaluate. The Predictor can then be used to request additional data from any source to aid in differentiation. We have explored this approach by analyzing the spectrum of a compound with the performance program. The Predictor was used to evaluate the the set of candidate structures to define the minimum number of metastable defocussing experiments necessary to achieve a unique solution. Thus, no time need be spent acquiring unnecessary or redundant data. Clearly, this has important implications for future work in that many different types of data (e.g., NMR, IR) might be requested by the Predictor to facilitate identification.

PLANS

For the remaining period of this grant we propose to carry out the following extensions of the research outlined above.

(I) PLANNER

The major area of activity related to the present version of the Planner will be to focus our attention on using the program in support of chemical studies outlined under Part B (see below). The chemical extraction and derivitization procedures used in the analysis of body fluids restricts the types of compounds present in each separated fraction. Such simplifications make this a problem more amenable to attack. Only certain classes of compounds are present in each fraction, and we have some knowledge of the mass spectral fragmentation of these classes. We wish to couple the program to the results of library matching procedures so that we direct our efforts to structure elucidation of those components which have not been previously identified. This is particularly important in the context of analysis problems such as those discussed under Part B.

We propose increasing the utility of the program by removing two present constraints: (a) allow unspecified "dummy" atoms in the skeleton instead of requiring a rigidly fixed structural skeleton, and (b) allow fragmentation processes to be specified more flexibly - in particular, allow fragmentations in substituents on the skeleton instead of requiring all fragmentations to cut through the skeleton.

Although we are presently uncomfortable with immediate coupling of the Structure Generator to the Planner, we propose continued exploration of the problems of controlling the generator automatically. Actual implementation awaits a more comprehensive treatment of the problem of constraints.

II. STRUCTURE GENERATOR

The inclusion of a reasonable set of constraints is obviously required and will be the subject of most of our future development work. We plan to develop an interface to the present interactive version of the Structure Generator that speaks a more chemical language. This interface will be designed to avoid the present requirement that the user know something about the program before he can use it. As the optimum method for implementation of a constraint is determined, the interface will be extended to translate the usual specification of the constraint in chemical terms into rules acting at the level of the program. As we stressed in development of the PLANNER, there are considerable advantages to building a powerful program in an

incremental fashion. These steps are logically directed to our longer term goal of developing a useful structure elucidation tool for the chemist, based on the structure generator.

There are several other areas of interest which are peripherally related to the problem of constraints and which will occupy our attention. The Structure Generator knows no chemistry other than atom names and their associated valence. There are several important areas where this is an immediate problem. For example, the program has no explicit awareness of the aromatic resonances, leading to a remediable redundancy in the list of isomers. An aromaticity-predictor is also indispensable for anticipating chemical behavior of a structure.

We wish to deal with types of isomers besides simple connectivity isomers. We need to have the facility for assembly of molecular sub-structures (the usual type of information inferred from spectroscopic data) when such an assembly yields new rings or multiple bonds. All the above questions need a reexamination of the fundamental mathematical considerations. The present algorithm has been proven to yield complete and irredundant solutions. In devising new algorithms or variants of the present one, the burden of proof can be reduced to (the usually easier) equivalence to the previous algorithm. Professor Harold Brown, who was the mathematician instrumental in initial development of the labelling algorithms for structure generation, will be with us again for several months to help attack the problems outlined above.

III. PREDICTOR

Although the Predictor has been essentially finished for our own internal use, we propose to spend a modest amount of time in the coming months making it more usable by others. In particular, we wish to extend the initial work on predicting the new experiments necessary for distinguishing among candidate structures (e.g., predicting that a metastable peak at mass 70.1 would confirm one structure and disconfirm another). In addition, we plan to work on cataloging some existing sets of mass spectrometry rules in such a way that the program can be easily used for different classes of problems.

Part A references (Published or submitted during year)

R. Carhart and C. Djerassi, "Applications of Artificial Intelligence for Chemical Inference XI....

D.H. Smith, B.G. Buchanan, R.S. Engelmore, H. Adlercreutz, and C. Djerassi, "Applications of Artificial Intelligence for Chemical Inference, IX. Analysis of Mixtures Without Prior Separation as Illustrated for Estrogens," J. Amer. Chem. Soc., September 5, 1973.

D.H. Smith, A.M. Duffield, and C. Djerassi, "Mass Spectrometry in Structural and Stereochemical Problems, CCXXII. Delineation of Competing Fragmentation Pathways of Complex Molecules from a Study of Metastable Ion Transitions of Deuterated Derivatives," Org. Mass Spectrom., 7, 367 (1973).

D.H. Smith, C. Djerassi, K.H. Maurer, and U. Rapp, "Mass Spectrometry in Structural and Stereochemical Problems, CCXXXIV. Applications of DADI, A Technique for Study of Metastable Ions, to Mixture Analysis," J. Amer. Chem. Soc., submitted (1973).

J.M. Block, D.H. Smith, and C. Djerassi, "Mass Spectrometry in Structural and Stereochemical Problems, CCXXXVIII. The Effect of Heteroatoms upon the Mass Spectrometric Fragmentation of Cyclohexanones," J. Org. Chem., submitted (1973).

L.M. Masinter, N.S. Sridharan, J. Lederberg, and D.H. Smith, "Applications of Artificial Intelligence for Chemical Inference XII. Exhaustive Generation of Cyclic and Acyclic Isomers," J. Amer. Chem. Soc., submitted (1973).

L.M. Masinter and N.S. Sridharan, "Applications of Artificial Intelligence for Chemical Inference, XIII. Labelling Objects Having Symmetry, J. Amer. Chem. Soc., submitted (1973).

D.H. Smith, L.M. Masinter, and N.S. Sridharan, "Heuristic DENDRAL: Analysis of Molecular Structures," to be published in the proceedings of the NATO/CRS Advanced Study Institute on Computer Representation and Manipulation of Chemical Information.

N.S. Sridharan, "Computer Generation of Vertex Graphs", Stanford Computer Science Memo CS-73-381, Stanford University, July, 1973.

Part B-i Gas Chromatograph - Mass Spectrometer Data System Development

OBJECTIVES AND RATIONALE

The objectives of this part of the research project are the improvement of GC/MS data system capabilities and the coupling of extracted data to the Heuristic DENDRAL programs for analysis. We ultimately seek a substantial degree of interaction between the instrumentation and the analysis programs including computer specification and control of the data to be collected. In addition to the development goals, this portion of the project provides for the day-to-day operation of the GC/MS systems in support of mass spectrum interpretation computer program development (Parts A and C) and applications of GC and MS to biomedical and natural product sample analysis with collaborators.

Our rationale for this approach is that the overall system should be designed for problem solving rather than just for data acquisition. This implies that analytical computer programs, after review of available experimental data, could be able to specify additional information needed to confirm a solution or distinguish between alternative solutions. Such requests could be passed back to an instrument management program to set up proper instrument parameters and collect the additional information. Our initial objectives to implement an on-line, closed-loop system using the ACME computer facility have met with a number of difficulties. These grow principally out of ACME's limited computing capacity and commitments as a general time-sharing service. In addition, the scanning high resolution mass spectrometer has inherent sensitivity limitations, which do not preclude a demonstration but rather limit the practical sample volume which could be analyzed. Until such limitations can be overcome, particularly in terms of computing support, we have focussed our efforts on an open-loop demonstration of such an approach.

PROGRESS

Progress has been made in demonstrating a GC/High Resolution Mass Spectrometry capability, in further developing automated data analysis algorithms, and in planning for the implementation of a data system for the collection of metastable ion information. Progress in these and other areas directed toward the main research goals has been impacted by a transition in computing support which is still underway. This transition, discussed in more detail below, was occasioned by the phase-over of the ACME computing facility, which we had been using, from NIH grant subsidy to a fully fee-for-service operation under Stanford University auspices.

Summaries of the results and problems encountered in each of the areas follow.

Gas Chromatography/High Resolution Mass Spectrometry (GC/HRMS)

We have verified the feasibility of combined gas chromatography/high resolution mass spectrometry (GC/HRMS). Using programs described in previous reports, we can acquire selected scans and reduce them automatically. The procedures are slow compared to "real-time" because of the limitations of the time-shared ACME facility. We have recorded sufficient spectra of standard compounds to show that the

system is performing well. A typical experiment which illustrates some of the parameters involved was the following. A mixture (approximately 1 microgram/component) of methyl palmitate and methyl stearate was analyzed by GC under conditions such that the GC peaks were well separated and of approximately 25 sec. duration. The mass spectrometer was scanned at a rate of 10.5 sec/decade, and a resolving power of 5000. The resulting mass spectra displayed peaks over a dynamic range of 100 to 1 and were automatically reduced to masses and elemental compositions without difficulty. Mass measurement accuracy appears to be 10 ppm over this dynamic range.

We have begun to exercise the GC/HRMS system on urine fractions containing significant components whose structures have not been elucidated on the basis of low resolution spectra alone. Whereas more work is required to establish system performance capabilities, two things have become clear: 1) GC/HRMS can be a useful analytical adjunct to our low resolution GC/MS clinical studies (Part B-ii), and 2) the sensitivity of the present system limits analysis to relatively intense GC peaks. This sensitivity limitation is inherent in scanning instruments where one gives up a factor of 20-50 in sensitivity over photographic image plane systems in return for on-line data read-out. This limitation may be relieved by using television read-out systems in conjunction with extended channeltron detector arrays as has been proposed by researchers at the Jet Propulsion Laboratory. We can nevertheless make progress in applying GC/HRMS techniques to accessible effluent peaks and can adapt the improved sensor capability when available.

These experiments have also shown that the ACME computer facility cannot reliably provide the rapid service required to acquire and file repetitive spectometer scans. This problem is to be expected in a heavily used time-shared facility without special configuration for high rate, real time support. Excepting possible requirements for real time data analysis (such as in a closed-loop system), this problem could be solved by implementing a large local buffer (e.g., disk) at the front-end data acquisition mini-computer. We are exploring this possibility in conjunction with the overall planning for computer support discussed below.

Data Analysis Algorithms

A. Peak Resolution

One of the significant trade-offs to be made in GC/HRMS is that of sensitivity versus resolution. In maintaining high instrument resolution (in the range of 5,000-10,000) while scanning fast enough to analyze a GC effluent peak (approximately 10 sec/decade), system sensitivity is constrained as discussed above. We have worked on a method for reducing instrument resolution requirements through more sophisticated computer analysis of a lower resolution output. In effect this transfers the burden of overlapping peak detection and mass determination to the computer instead of requiring inherently well resolved data out of the instrument. The advantage comes in better system sensitivity.

Unresolved peaks are separated by an analytical algorithm, the operation of which is based on a model peak derived from known singlet peaks in the data. Actual tabulated peak models are used rather than the assumption of a particular parametric shape (e.g., triangular, Gaussian, etc.). This algorithm provides an effective

increase in system resolution by approximately a factor of three thereby effectively increasing system sensitivity. By measuring and comparing successive moments of the sample and model peaks, a series of hypotheses are tested to establish the multiplicity of the peak, minimizing computing requirements for the usually encountered simple peaks. Analytic expressions for the amplitudes and positions of component peaks have been derived in the doublet case in terms of the first four moments of the peak complex. This eliminates time consuming iteration procedures for this important multiplet case. Iteration is still required for more complex multiplets.

B. GC Analysis

The application of GC/MS techniques to clinical problems as described in Part B(ii) of this proposal has indicated the desirability of automating the analysis of the results of a GC/MS experiment. The GC/MS output involves extracting from the approximately 700 spectra collected during a GC run, the 50 or so representing components of the body fluid sample. The raw spectra are in part contaminated with background "column bleed" and in part composited with adjacent constituent spectra unresolved by the GC.

We have begun to develop a solution to this problem with promising results. By using a disk-oriented matrix transposition algorithm, the array of 700 spectra by 500 mass samples per spectrum can be rotated to gain convenient access to the "mass fragmentogram" form of the data. The transposition algorithm avoids many successive passes over the input data file as would be required in a straightforward approach. By generating a reorganized intermediate file, time savings by factors of 5-10 are achieved.

The fragmentogram form of the data displayed at a few selected mass values, has been used at Stanford, MIT, and elsewhere for some time to evaluate the GC effluent profile as seen from these masses. Mass fragmentograms have the important property of displaying higher resolution in localizing GC effluent constituents. Thus by transposing the raw data to the mass fragmentogram domain, we can systematically analyze these data for baselines, peak positions, and amplitudes, and thus derive better mass spectra for the relatively few constituent materials. These are free from background contamination and influences of adjacent GC peaks unresolved in the overall gas chromatogram. These spectra can then be analyzed by library search techniques or first principles as necessary.

We have applied a preliminary version of this algorithm to several urine samples. These contain several apparently simple peaks which in fact consist of multiple components. The algorithm performs well in separating out these constituents although further testing is required.

Closed-loop Instrument Control

In the long term, it could be possible for the data interpretation software to direct the acquisition of data in order to minimize ambiguities in problem solutions and to optimize system efficiency. The task of deciding among and collecting various types of mass spectral information (e.g., high resolution spectra, low ionizing voltage spectra, or selected metastable ion information) under closed-loop control during a GC experiment is difficult. Problems arise because of the large requirements placed on computer resources

and present limitations in instrument sensitivity or data read-out imposed by the time constraints of GC effluent peak widths. Solutions to these problems may not be economically feasible within currently existing technology but seem achievable in the future. We are studying this problem in a manner which would entail a multi (two or three) - pass system. This permits the collection of one type of data (e.g., high resolucion mass spectra) during the first GC/MS analysis. Processing of these data by DENDRAL will reveal what additional data are necessary on specific GC peaks during a subsequent GC/MS run. Such additional data could help to uniquely solve a structure or at least to reduce the number of candidate structures. This simulated closed-loop procedure could demonstrate the utility of DENDRAL type programs to examine data, determine solutions and propose additional strategies, but will not have the requirement of operating in real-time. Some parameters in the acquisition of particular types of information, such as metastable data, will require computer control, even in the open-loop mode.

We have considered plans to implement two aspects of instrument control, in addition to the magnetic scan control implemented for GC operation and reported previously. These include system resolution control, such as would be required to change from normal spectrum scanning mode to metastable scanning mode, and high voltage control necessary to selectively measure metastable ion fragmentation data. In addition to these we have considered the discrete switching of various electronic mode controls which are straightforward and not discussed in detail.

Implementation plans for computer control of these instrument functions have been delayed because of the ACME computing facility transition which diverted the necessary hardware and software manpower.

Resolution control involves changing the widths of the slits at the exit of the ion source and the entrance to the ion multiplier detector. Additional source and electrostatic analyzer voltages must be controlled to optimize performance, as discussed later. Mechanical slit adjustment is accomplished on the MAT-711 instrument by heating wires which support the slit jaws. The resulting expansion or contraction of those wires move the spring-loaded jaws. As implemented by the manufacturer, the time constants involved in heating the control wires are 5-10 seconds. It is possible to speed this up to approximately 0.5 seconds by application of a controlled over voltage decreasing to the appropriate equilibrium value for the desired slit width. This was demonstrated by a series of experiments on an extra slit assembly mounted in a vacuum jar in our laboratory. Cooling of the wires is relatively fast in the way they are mounted so no problem exists in that direction.

It is desirable to have feedback to indicate the actual slit width achieved rather than relying on a slit assembly calibration. Stretching of the support wires or changes in the spring tension under temperature cycling would change this calibration. An optical scheme to measure slit width in situ is possible. We do not contemplate implementing this feedback immediately because it requires major changes to the instrument flight tube.

Two types of metastable ion relationships are obtainable by suitable control of the double-focussing instrument. First, for a given daughter ion, one can trace the parent ions which give rise to

it. Second, for a given parent ion one can trace the various daughters to which it gives rise. The first measurement ("metastable defocussing") is the more straightforward for this instrument since parent ions can be enumerated by a simple scan of the accelerating voltage, holding the electrostatic analyzer (ESA) voltage and magnetic field constant. The second type of scan requires the coordinated scan to two of the three fields. We feel that joint computer control of the accelerating voltage and ESA voltage is the simpler approach since the magnetic field is more difficult to set and monitor because of hysteresis effects. For a resolution of 1000 in the metastable ion mass measurement, the voltages must be set to approximately .01-.02% accuracy. This requires a 14-16 bit digital-to-analog (D/A) converter to control the input (10 volts) to the operational amplifier which generates the high voltage. Similar D/A controls of ion source voltages for ion current and focus optimization can be implemented using optical isolators to allow vernier control of the various high voltages around the nominal 8KV values.

Computing Transition

As mentioned earlier, the transition of the ACME computing facility from NIH subsidy to Stanford-sponsored fee-for-service operation has impacted our development efforts this past year. Both the low resolution instrument used for routine body fluid analysis research and the high resolution instrument are affected. All computing support was previously obtained from the ACME facility, much of it as core research without explicit transfer of funds. The transition has required consideration of both technical and economic factors. The new facility represents a combination of the previous ACME interactive and real time computing load with various administrative and batch computing loads on a new IBM 370/158 computer. This new environment will have even more difficulty in supporting real time computing needs than ACME did. No real time support has been available since the 360/50 service was discontinued on July 31, although terminal service was reestablished in mid-August. Data acquisition service via the IBM 1800 is expected to be operational by early November.

For the high resolution instrument, this transition, as a minimum, necessitates an interface modification (we previously sent data through the IBM 2701 interface no longer to be supported). It also amplifies the problems we encountered in sending and filing high rate mass spectrometer data (particularly during GC/MS runs). These problems would be present to some extent in any general time-sharing service machine without specific hardware and software configuration provision for these needs (such provisions for real time support had been proposed in our SUMEX computer application).

After examining a variety of alternatives, we conclude that a dedicated mini-computer solution (built around a machine with the arithmetic capability of a PDP-11/45) would be highly attractive technically and relatively inexpensive. A stand-alone mini-computer system would cost in the range of $50,000-$60,000, augmenting existing equipment, plus approximately $9,000 per year maintenance and $2,000 per year for supplies. Estimates for 370/158 support, based on current charging algorithms and previous utilization experience, run from $35,000-$50,000 per year. This spread is caused by uncertainties in the effects of planned measures to increase operating efficiency and possible changes to the rate structure. In

any case, the mini-computer approach pays for itself in 1 to 2 years of operation and provides the responsiveness of a dedicated machine for real time support. Unfortunately our existing budget does not provide for this solution. The budget is very marginal for purchase of computing support from the 370/158 as well. This later approach is the only currently available one, however, since it can be implemented with relatively low start-up cost. The effect of budget limitations appears in terms of a reduced number of samples which can be run. We have attempted to minimize the other budget costs (manpower principally) to increase the computing funds available. This will necessarily impact our development goals. We hope, in the renewal application for DENDRAL support, to be able to implement the more effective mini-computer approach for the high resolution spectrometer as a longer term solution.

We have undertaken an interim mini-computer solution for the low resolution spectrometer (Finnigan 1015 quadrupole) which is primarily used for our body fluid analysis studies. For the same reasons outlined above, a mini-computer solution is attractive. In the case of the low resolution quadrupole instrument, a lesser capacity machine will suffice for immediate data acquisition and display functions. We have implemented such an interim system on a PDP-11/20 machine available from other funding sources. This system, which is now operational, allows the acquisition of GC/MS data, limited by the capacity of the DEC tape storage medium to approximately 600 spectra, per experiment. For certain types of GC analyses, up to 1000 spectra per experiment are required so this limits, to some extent, the utility of this interim system. A calcomp plotter is supported for display purposes. A fixed head disk provides for library search procedures which are still being converted from the ACME system. We have applied to the NIH-GMS for funds to augment this system in order to relieve current limitations as part of a Genetics Center research proposal.

FUTURE PLANS

Our future plans are basically to continue development along the lines outlined above. We will complete the computing support transition steps described. These include primarily establishing a connection to the new 370/158 facility to provide interim support for the high resolution system. We will pursue additional software and hardware development goals as far as possible within the limited budget available. These efforts will concentrate for the most part on bringing up a metastable ion analysis data system. It should be reemphasized that the manpower levels proposed in the follow-on budget have been minimized to allow for purchasing computing time on the 370/158. The allocated manpower is required primarily for instrument operation and maintenance with minimal provision for development efforts.

Part B(ii). Analysis of Body Fluids by Gas Chromatography/Mass Spectrometry.

The chemical separation of urine into the following fractions prior to GC/MS analysis has been described in previous DENDRAL Reports:

free acids (analyzed by gc/ms as their methyl esters)
amino acids (analyzed by gc/ms as their N-trifluoroacetate n-butyl ester)
carbohydrates (analyzed by gc/ms as their trimethyl silyl ether derivatives)
hydrolyzed acids (analyzed by gc/ms as their methyl esters)
hydrolyzed amino acids (analyzed by gc/ms as their N-trifluoroacetate n-butyl esters)

During the past year we have extended these methods of fractionation to the following body fluids: blood (after an initial precipitation of proteins by the addition of ethanol) and amniotic and cerebrospinal fluids. The following summarizes the results obtained from an analysis of these fluids during the past year by gas chromatography-mass spectrometry.

URINE ANALYSIS:

A. The Development of a "Metabolic" Profile Characteristic of Neonatal Tyrosenemia Using Combined Gas Chromatography-Mass Spectrometry.

This work was carried out in collaboration with clinical colleagues from the Department of Pediatrics at Stanford University and a joint publication describing this research is in preparation.

The study was based on a total of one hundred and four 24-hour urine samples from sixteen premature or small birthweight infants receiving treatment in the Stanford nursery. After exclusion of infants who became ill, died, or left the nursery, we were able to follow nine infants closely for periods of between 4 and 6 weeks from day 3 of life. All nine infants had birthweights of below 1500g and three of these were below 1000g.

Of the nine infants studied, five showed transient tyrosinemia as shown by a marked elevation in the urinary excretion of the tyrosine metabolites, p-hydroxyphenyllactic acid, p-hydroxyphenylpyruvic acid and p-hydroxyphenylacetic acid. There was also a less marked but distinct elevation in the urinary tyrosine output. Figures 1 and 2 show the metabolic profiles of the same infant (J.L.) in the normal(a) and tyrosinemic(b) states. Figure 1 shows the free acid outputs, chromatographed as the methyl ester-methyl ether derivatives and Figure 2 is an expression of the free amino acids of the same urines, chromatographed as the N-trifluoroacetyl n-butyl ester derivatives. In each case the concentration of each metabolite is a function of the peak height as compared to the height of the internal standard. Table 1 is a summary of the ranges of urinary output of tyrosine and metabolites observed for all the infants in the study.

TABLE 1 Daily Excretion in mg/kg

Tyrosine	p-HPLactic	p-HPPyruvic	p-HPAcetic

Normal	0.2 - 3	0 - 5	0 - 0.5	0.2 - 2
Tyrosinemic	3 - 15	5 - 50	0.5 - 5	0.5 - 5

As shown by Table 1 and Figure 1 neonatal tyrosinemia is characterized by a very large increase in the output of p-hydroxyphenyllactic acid and by a 10-50 fold excess of the latter over p-hydroxyphenylpyruvic acid. Studies of the hereditary defects in tyrosine metabolism initially indicated that p-hydroxyphenylpyruvic acid was the major metabolite although more recently cases have been reported where p-hydroxyphenyllactic is in a 2-5 excess over p-hydroxyphenylpyruvic. These latter determinations were made using GC and GC/MS methods and therefore probably reflect the improved specificity of the analytical procedure (previously colormetric methods were used) rather than a difference of actual metabolic profile. Apart from the very large excess of p-hydroxyphenyllactic acid over its keto analog we could detect no significant differences between the profiles shown in neonatal tyrosinemia and those published for hereditary disease. Other metabolites such as p-hydroxymandelic acid, DOPA N-acetyltyrosine, which have previously been reported in tyrosinemic urine were not seen to be elevated.

B. GC/MS Analysis of Urine from Children Suffering from Leukemia.

This research was carried out with twenty 24-hour urine samples supplied by Drs. Jordan Wilbur and Tom Long of the Stanford Children's Hospital.

The acidic fraction of all urines studied in this project showed no abnormal metabolites nor were gross amounts of known acids detected. The amino acid fraction, however, of six of the urine samples showed the presence of an non-protein amino acid, beta-aminoisobutyric acid (BAIB). In several of these instances the patients were excreting in excess of 1 gram of BAIB per day. The literature contains many references to increased BAIB excretion (genetic excretors, lead poisoning, pulmonary tuberculosis, march hemoglobinuria, thalassaemia and Down's Syndrome). The reported excretion of BAIB by leukemic patients was not substantiated by another investigator. There are several criticisms in the literature of the methods used for the quantitation of BAIB in biological fluids and in order to fill this void a sensitive, specific and rapid method for the quantitation of BAIB has been developed. (SEE: The Quantitation of bAIB in Urine by Mass Fragmentography; W.E. Pereira, R. Summons, W.E. Reynolds, T.C. Rindfleisch and A.M. Duffield, in press).

C. GC/MS Analysis of Urine from Patients Suffering from Hodgkin's Disease.

During this study 20 urine samples from patients with diagnosed Hodgkin's Disease (Department of Oncology, Stanford University Medical Center) were analyzed and in general, no abnormal metabolic profile could be found in any urine. There was one exception in which an individual was noted to excrete massive quantities of adipic acid (of the order of 1 gram per day).

D. Detection of Metabolic Errors by GC/MS Analysis of Body Fluids.

This project results from a collaborative effort between the Departments of Genetics and Pediatrics of the Stanford University Medical Center. To date over 50 samples have been analyzed; the majority (35) being

urine, while amniotic fluid (10), blood (6), and cerebrospinal fluid (6) were also analyzed. It has been and will continue to be our practice to analyze aliquots of fluid samples in collaboration with clinical investigators obtained for valid diagnostic purposes completely divorced from this research on GC/MS analysis techniques. This investigation is not intended to serve as a screening program for a large population but rather to focus on those individuals who exhibit suggestive clinical manifestations such as psychomotor retardation and progressive neurologic disease as well as suggestive pedigrees.

In the case of amniotic fluid the hope is to be able to monitor the condition of the fetus in those pregnancies which might be considered at risk. To date we have investigated specimens from normal pregnancies in order to establish the catalog of compounds to be observed in amniotic fluid. From this base it could prove possible to identify materials which might identify the health of the fetus.

We have been able to confirm the presence of orotic acid in a urine from a person found to have orotic aciduria while another urine sample was used to demonstrate our ability to identify the characteristic metabolites present in isovaleric acidemia. The following description refers to a urine from a child with hypophosphatasia.

A child died 33 hours after birth in Fresno, California, with the classical signs of hypophosphatasia. This genetic defect is marked by high phosphoethanolamine (PEA) concentrations in urine of affected homozygotes and unaffected heterozygotes. After derivatization (in this instance the TMS ethers of the water soluble carbohydrate fraction were prepared) we were able to detect by GC/MS large concentrations of ethanolamine and phosphoric acid but not PEA itself. The derivatization procedure we used most likely hydrolyzed PEA. We were able to quantitate for this compound in the infant's urine using an amino acid analyzer, and PEA excretion was extremely high (over 200 times normal values for infants) confirming the diagnosis. Next we examined urine samples from the child's parents, presumed heterozygotes, by GC/MS and by the amino acid analyzer. Again, no PEA was detected by the former method although the presence of ethanolamine and phosphoric acid was demonstrated. We determined the following excretion levels of PEA by amino acid analyzer:

Newborn infant:	94 micromoles per 100 ml.	(Normal 0.21-0.33)
Father:	269 micromoles per 24 hours	(normal 17-99)
Mother:	32 micromoles per 24 hours	(normal 17-99)

It is of interest that in this family the affected infant and his unaffected father both show subnormal serum alkaline phosphatase activity. The mother, who did not excrete increased amounts of PEA, was found to have normal activity of this enzyme in her serum. The following table summarizes the serum phosphatase activity measurements:

Newborn infant:	0.2 units*	(normal 2.8-6.7)
Father:	0.7 units	(normal 0.8-2.3)
Mother:	3.4 units	(normal 0.8-2.3)

(* - 1 unit is that phosphatase activity which will liberate 1 millimole of p-nitrophenol per hour per liter of serum)

E. Drug Analysis Service Using GC/MS

We were recently contacted by physicians to rapidly identify a drug self-administered by a patient in the Stanford University Hospital. From the mass spectrum the drug was identified as pentazocaine within the hour. Although not part of the formal DENDRAL proposal we expect that similar cases may arise in the future and we intend to respond positively to such requests.

Development of Library Search Routines for Mass Spectrum Identification

The analysis of a single body fluid fraction produces between 600 and 750 mass spectra. In order to cope with the interpretation of the daily production of mass specta (about 8 body fluid fractions for a total of between 4,800 and 6,000 mass spectra) we have begun the implementation of library search routines. Concurrent with the analysis of body fluids for metabolic content we have been recording the mass spectra of many reference compounds. This collection represents the beginning of the construction of a library of reference spectra. Late in 1973 we expect to receive from Dr. S. Markey, University of Colorado Medical Center, a more comprehensive library which he has collated from contributors (including our own laboratory) in the field of biological applications of gas chromatography/mass spectrometry.

Prior to the demise of the ACME computer faciliity at Stanford University, we ran library search routines on data collected from urine fractions. Because of the ACME system being heavily loaded, our programs took about one minute per compound identification. However, the experience gained will be used to implement library search routines on our current PDP-11 GC/MS data system. In addition we have sent mass spectra from several urine analyses to Dr. S. Grotch, Jet Propulsion Laboratory, Pasadena, California, in order that he could use his library search routines on real data. In this instance the limiting factor for efficient compound identification was the library content which was limited to a few compounds of biological significance. In addition those compounds of interest that were present in the library were often in a derivatized form different from that used in our analytical methodology.

Application of GC/HRMS to Body Fluid Analysis

We reported in the last annual report of the DENDRAL project that the Varian MAT 711 mass spectrometer was interfaced with a gas chromatograph for the recording of low resolution mass spectra. We have now used this system for the recording of HRMS of gas chromatographic fractions from urine analyses. We were able to record HRMS scans over several gas chromatographic peaks of interest in a number of urine fractions. The high resolution results were found to be of a high quality in mass measurement accuracy. When using the MAT 711 instrument for GC/HRMS the sensitivity of the ion source was a limiting factor in that less intense gas chromatographic peaks often lacked sufficient material to generate acceptable high resolution mass spectra. Notwithstanding this limitation the HRMS data recorded on different urine fractions was used to confirm the identification of several metabolites. If by chance the metabolite of concern was available only in quantities insufficient for direct GC/HRMS, preparative GC would be used to concentrate the component of interest for subsequent HRMS.

RESOURCE OPERATION

Over the term of this grant our mass spectrometry laboratories have provided support to numerous research projects in addition to the DENDRAL computer program development project funded under this grant. These cover a variety of applications at Stanford, in the United States, and abroad. Included are problems in the study of human metabolites, biochemistry, and natural product chemistry. Samples have been run in collaboration with outside people both on the MAT-711 GC/High Resolution Mass Spectrometer system and the Finnigan 1015 GC/Low Resolution Quadrupole Mass Spectrometer system. The low resolution system has also been supported by a NASA research grant.

The following tables summarize the support rendered in terms of numbers of samples run through various types of analysis:

I. MAT-711 High Resolution System (Period covered 11/71 - 6/73).

	Batch High Resol. MS	Batch Low Resol. MS	GC/High Resol. MS	GC/Low Resol. MS
DENDRAL program devel.	317	3		
Stanford Genetics (Body fluid analysis)	39	17	13	
Stanford Chemistry (non-DENDRAL - Dr. Djerassi's group)	91	112		50
Stanford Chemistry (non-DENDRAL - Drs. Vantamelen, Johnson, Mosher, Collman, Altman, Goldstein)	29	23		4
Stanford Surgery (Dr. Fair)	8			
Dr. Adlerkreutz (Finland)	10			
Dr. Venien (France)	26			
Dr. Gilbert, Mors, Baker (Brazil)	40	44		
Dr. Orazi (Argentina)	19	1		
Dr. Subramanian (India)	10	5		
Dr. Khastgir (India)		5		
Dr. O'Sullivan (Ireland)		5		
Dr. Badr (Libya)	30			
Dr. Mital (India)	5			

624 samples	215 samples	13 samples	54 samples

II) FINNIGAN 1015 Low Resolution System (period covered 8/72-8/73)

Note the samples run are specified by fluid type. Each fluid is extracted and derivatized as described in Part B (ii) and therefore may represent several GC/LRMS analyses. Specific discussions of the results of various of the analyses run are discussed earlier in Part B(ii).

	GC/Low Resolution MS
Stanford Pediatrics (Drs. Cann, Sunshine and Johnson)	141 urines
	7 Amniotic Fluids
	6 bloods
	2 cerebrospinal fluids
Stanford Oncology (Dr. Rosenberg)	20 urines
Stanford Psychiatry - Genetics (Drs. Brodie and Cavalli-Sforza)	4 cerebrospinal fluids
Stanford Respiratory Medicine (Dr. Robin)	2 urines
	2 bloods
Stanford Pharmacology (Dr. Kalman)	2 extracts
Stanford Biochemistry (Dr. Stark)	4 extracts
Stanford Children's Hospital (Drs. Wilbur and Long)	24 urines
UC San Francisco Medical School - Dermatology (Dr. Banda)	2 urines
Menlo Park V.A. Hospital (Dr. Forrest)	13 extracts
Palo Alto V.A. Hospital (Drs. Hollister and Green)	7 extracts
University of Puerto Rico School of Medicine (Dr. Garcia-Castro)	7 urines
	243 samples

PART B PUBLICATIONS

The following summarizes the publications resulting from research in the low resolution mass spectrometry laboratory over the past year, including body fluid analysis. This laboratory has been jointly supported by NIH (DENDRAL) and NASA. The listed publications include research relevant to both sponsors.

The Determination of Phenylalanine in Serum by Mass Fragmentography.
Clinical Biochem., 6 (1973)
By W.E. Pereira, V.A. Bacon, Y. Hoyano, R. Summons and A.M. Duffield.

The Simultaneous Quantitation of Ten Amino Acids in Soil Extracts by Mass Fragmentography
Anal. Biochem., 55, 236 (1973)
By. W.E. Pereira, Y. Hoyano, W.E. Reynolds, R.E. Summons and A.M. Duffield.

An Analysis of Twelve Amino Acids in Biological Fluids by Mass Fragmentography.
Anal. Chem.,
By R.E. Summons, W.E. Pereira, W.E. Reynolds, T.C. Rindfleisch and A.M. Duffield.

The Quantitation of B-Amino isobutyric Acid in Urine by Mass Fragmentography.
Clin. Chim. Acta, in press
By W.E. Pereira, R.E. Summons, W.E. Reynolds, T.C. Rindfleisch and A.M. Duffield.

The Determination of Ethanol in Blood and Urine by Mass Fragmentography.
Clin. Chim. Acta
By W.E. Pereira, R.E. Summons, T.C. Rindfleisch and A.M. Duffield.

A Study of the Electron Impact Fragmentation of Promazine Sulphoxide and Promazine using Specifically Deuterated Analogues.
Austral. J. Chem., 26, 325 (1973)
By M.D. Solomon, R. Summons, W. Pereira and A.M. Duffield.

Mass Spectrometry in Structural and Stereochemical Problems. CCXXXVII. Electron Impact Induced Hydrogen Losses and Migrations in Some Aromatic Amides
Org. Mass Spectry., in press.
By A.M. Duffield, G. DeMartino and C. Djerassi.

Spectrometrie de Masse. IX. Fragmentations Induites par Impact Electronique de Glycols- -En Serie Tetraline
Bull Soc. Chim. France, 2105 (1973).

Spectrometric de Masse VIII. Elimination d'can Induite par Impact Electronique dans Le Tetrahydro-1,2,3,4-Napthtal-ene-diol-1,2.
Org. Mass Spectre., 7, 357 (1973).
By P. Perros, J.P. Morizur, J. Kossanyi and A.M. Duffield.

Chlorination Studies I. The Reaction of Aqueous Hypochlorous Acid with Cytosine.
Biochem. Biophys. Res. Commun., 48, 880 (1972)
By W. Patton, V. Bacon, A.M. Duffield, B. Halpern, Y. Hoyano, W. Pereira and J. Lederberg.

Chlorination Studies II. The Reaction of Aqueous Hypochlorous Acid with -Amino Acids and Dipeptides.
Biochim. et Biophys. Acta, 313, 170 (1973).
By. W.E. Pereira, Y. Hoyano, R. Summons, V.A. Bacon and A.M. Duffield.

Chlorination Studies IV. The Reaction of Aqueous Hypochlorous Acid with Pyrimidine and Purine Bases.
Biochem. Biophys. Res. Commun., 53, 1195 (1973).
By Y. Hoyano, V. Bacon, R.E. Summons, W.E. Pereira, B. Halpern and A.M. Duffield.

Part C. EXTENSION OF THE THEORY OF MASS SPECTROMETRY BY COMPUTER

OBJECTIVES:

Part C of the DENDRAL effort, termed Meta-DENDRAL, aims at providing theory formation help for chemists interested in the mass spectrometric behavior of new classes of compounds. Our goals are necessarily long-range because theory formation by computer is itself an exciting, unsolved problem in computer science. We have chosen to explore this problem in the context of mass spectrometry in order to make frontier computer research results available to working scientists.

The problem of finding judgmental rules for use in a computer program is common to many biomedical computing projects, such as medical diagnosis and therapy recommendation programs. <See, for example, Shortliffe, et.al.> In order to give these programs the knowledge that makes them perform at acceptable levels, a medical expert is often asked to summarize his own knowledge of the problem area in rules that the program can use. The Meta-DENDRAL theory formation program is a paradigm for the kind of assistance that computers can give to the medical experts in this role. Programs of this sort can, first of all, provide the expert with an interpreted summary of a large collection of "hard" empirical data. Second, the program can suggest to the expert plausible rules that appear to explain major features of the data. Thus, the expert is able to assimilate large collections of data in the rules given to the computer. We believe that the meta-DENDRAL work is a useful model on which fruitful work in other biomedical problems can be based.

The over-all strategy of this research is to model the theory formation activity of scientists. We start with a set of empirical data which are known molecular structures and their associated mass spectra. By exploring the possible mechanistic explanations of each mass spectrum, the program is able to find a set of mechanisms that appear to be characteristic for the class of molecules. These characteristic mechanisms constitute the general mass spectrometry rules for the class, or a first-level theory for the class. Further refinements of the rules give more sophisticated restatements of the theory.

We have designed the programs in such a way as to provide useful results from the intermediate steps. The progress section discusses several sets of results that have been obtained, even though the entire program has not yet been completed.

PROGRESS:

In the past ten months (since January, 1973) the theory formation programs have seen significant application and significant new extensions. In addition, the work has been described in publications for both chemists and computer scientists.

Applications of Existing Programs.

The INTSUM program, for interpreting and summarizing the mass spectra of many known compounds of one class, was described in the previous annual report as essentially finished. In this last period we have used this program to help understand the mass spectrometry of several

classes of compounds, including estrogens, equilenins and other estrogenic steroids, androstanes, alkyl pregnanes, vinyl quinazalones, amino acids and aromatic acids. An article written for mass spectroscopists and soon to appear in Tetrahedron (Smith, et.al, enclosed) describes this program and its usefulness in understanding the previously unreported mass spectrometry of the equilenins. The amino acid and aromatic acid results are useful for interpreting the mass spectra taken from those fractions of urine (see Part B).

The INTSUM program is available to anyone who requests it, as stated in the article soon to appear. Because of the complexity of the program, we recommend that mass spectroscopists use this program on a network computer after they have collected a number of mass spectra from a class of compounds whose fragmentation mechanisms they wish to investigate.

Recent Extensions to Meta-DENDRAL.

In this last period significant progress has been made on the theory formation programs that use the interpreted summary of the data provided by the INTSUM program. A simple rule formation program, described previously (MI7), finds the characteristic mass spectrometry mechanisms for a class of compounds, assuming that the compounds exhibit regular behavior as a class. Recent work has removed the restriction that the compounds must behave as a class - important classes can be found by the program within the set of given compounds. The procedure was described in a paper for the Third International Joint Conference on Artificial Intelligence, which is enclosed. At the same time that the rule formation program looks for characteristic mechanisms, the class separation procedure refines the class of molecules that appear to behave uniformly (i.e., appear to exhibit most of the characteristic mechanisms).

Another important extension of the theory formation program makes the rule descriptions more general and less specific to the class of compounds studied. The mechanisms in the rules are now described generally in terms of the kinds of bonds that break, and not in terms of the precise relations of the bonds to the skeletal structure common to the class. For example, a rule is now stated as "Any bond that is the second bond from a nitrogen atom is likely to break", rather than "In the skeleton R1-C2-N3-C4-R5 the bond between atoms 1 and 2 and the bond between atoms 4 and 5 are both likely to break".

These general descriptions will allow much more freedom in the kinds of interpretations that can be placed on the INTSUM results. It is possible, for example, to alter the set of predicates used to describe bonds without altering the program.

The program can be conceptualized as a search program through the space of possible combinations of predicates. Some predicates describe the type of bond (e.g., 'single'), others describe the atoms joined by the bond (e.g., 'nitrogen', 'secondary'), and others describe the bonds and atoms next away from the bond that breaks. Some a priori heuristics limit consideration of complex predicates to chemically meaningful combinations, for example, by forbidding consideration of a single atom as both carbon and nitrogen. Other heuristics guide the process of expansion by forbidding a new predicate to be added to a description if its addition reduces the explanatory power of the existing description. For example, if a high average intensity is associated with breaking the

X-X bond in X-X-N and further specification of either of the X's reduces the average intensity, then the description is not changed.

In addition to the work just mentioned, a generative model of rule formation has been pursued by Carl Farrell in his dissertation work directed by Professor Feigenbaum and Dr. Buchanan. He has written a program which accepts, as input, descriptions of specific molecules and all the primitive actions that might explain the mass spectra of those molecules. The output of the program is a set of general situation-action rules that describe classes of molecules that seem to be characteristically show evidence of significant actions.

PLANS

In the following period we plan to increase the performance capabilities of the theory formation program in several ways.

1. Sample Selection.

The program's current strategy is to find the rules exhibited by most or all of the molecules in the initial sample. If the molecules are diverse, the rules will be diverse. Thus, we plan to add a preprocessor that can select a "simple" set of molecules for the rule formation to work with. For example, unbranched (straight-chain) compounds should be expected to present fewer complications for initial theory formation than highly branched compounds. The effects of the complicating features can be studied after the simple rules have been found.

2. Rule Clarification.

After simple rules have been found, we want the program to clarify the conditions under which the rules hold. By studying more complicated molecules, the program can find the simple rules that no longer hold for these cases. For example, we want the program to discover that terminal alpha carbons (as in CH3-X-N) are special. Or, the program should discover the effects of double bonds by examining new cases even though the molecules in the original set contained no double bonds.

3. Experimentation.

Because the original set of molecules contains the simpler examples from which it is easier to find characteristic mechanisms, the program will need to clarify rules in the way suggested under (2). For a human scientist, this means describing new experiments to perform that will help place limits on the range of applicability of the rules. Looking at additional arbitrary molecules may be helpful, but not as helpful as looking at the specific molecules that will resolve specific questions about the preliminary rule set.

4. Integration of Results.

When the program has examined two or more classes of molecules, it should be able to integrate the results into a common set of mechanisms (if any are common). The set of predicates used by the integration program may not have to be wider than the set used by the rule formation program, but one would expect the rules themselves to be more general. For example, integrating aliphatic amine and ether results should combine the separate alpha-cleavage rules (one with nitrogen, one

with oxygen) into a more general rule (specifying 'N or O', or 'heteroatom').

PART C REFERENCES (Published or submitted during this year)

D.H. Smith, B.G. Buchanan, W.C. White, E.A. Feigenbaum, C. Djerassi and J. Lederberg, "Applications of Artificial Intelligence for Chemical Inference X. INTSUM. A Data Interpretation Program as Applied to the Collected Mass Spectra of Estrogenic Steroids". Tetrahedron. In press.

B.G. Buchanan and N.S. Sridharan, "Analysis of Behavior of Chemical Molecules: Rule Formation on Non-Homogeneous Classes of Objects". In proceedings of the Third International Joint Conference on Artificial Intelligence, Stanford University (August, 1973). (Also Stanford Artificial Intelligence Project Memo No. 215.)

Related Publications

D. Michie and B.G. Buchanan, "Current Status of the Heuristic DENDRAL Program for Applying Artificial Intelligence to the Interpretation of Mass Spectra". August, 1973.

E.H. Shortliffe, S.G. Axline, B.G. Buchanan, T.C. Merrigan and S.N. Cohen, "An Artificial Intelligence Program to Advise Physicians Regarding Antimicrobial Therapy". Computers & Biomedical Research. In Press.

HUMAN SUBJECTS

HUMAN SUBJECTS

As a part of this research project, GC/MS analysis techniques will be applied to human body fluids in collaboration with clinical investigators, and blood and urine specimens will be collected from human subjects. Collection of VOIDED URINE SPECIMENS presents no risk to the patient. Collection of blood samples will not be taken solely for the purpose of this research but rather would be collected as part of a diagnostic procedure deemed necessary for clinical diagnosis.

The undersigned agrees to accept responsibility for the scientific and technical conduct of the project and for provision of required progress reports if a grant is awarded as the result of this application.

12/19/73
Date

Carl Djerassi
Principal Investigator

APPENDIX A

FIGURES 1-3

To Josh
From Dennis

Here is the additional background material you requested on ~~the~~ multiplet resolution and associated topics

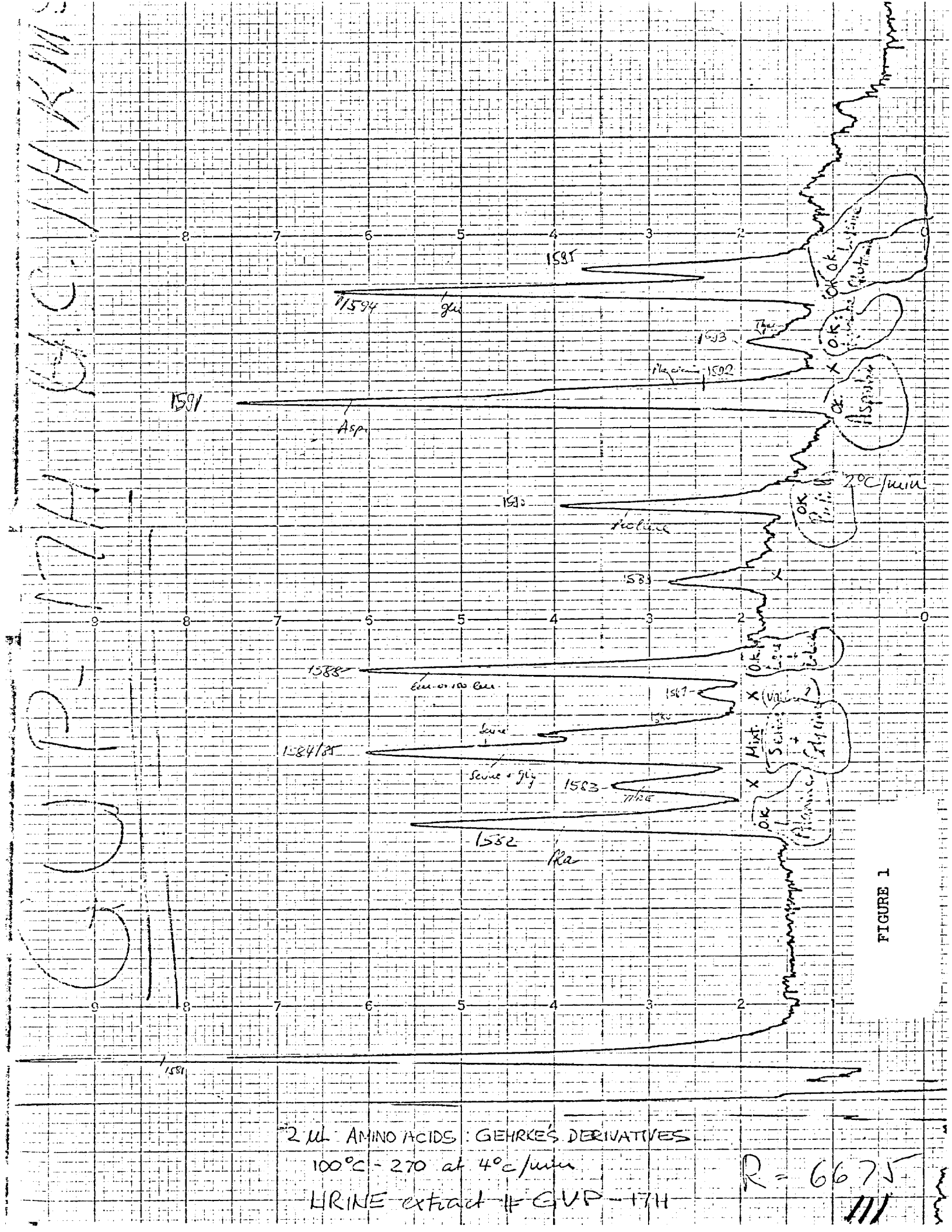
Dennis

do Statistics of Am

Table 1. Potential doublets in composition range C,H, + ... with mass differences (e) <50 mmu (0.05 amu).

C	H	N	O	E mmu	
0	0	0	0	0.0	
0	[illegible]	1	-1	23.4	..
0	[illegible]	2	-2	46.8	
0	-8	4	-3	-33.4	
1	-2	-3	2	-34.6	
1	0	-2	1	-11.2	..
1	2	-1	0	11.2	..
1	[illegible]	0	-1	35.6	
1	-6	2	-2	-44.7	
1	-6	3	-3	-21.3	
1	-4	4	-4	2.1	--
2	0	-4	2	-22.5	..
2	2	-3	1	0.9	--
2	4	-2	0	24.4	..
2	6	-1	-1	47.8	
2	-6	1	-2	-32.[illegible]	
2	-4	2	-3	-9.[illegible]	--
2	-2	3	-4	14.3	..
3	[illegible]	-4	1	13.1	..
3	6	-3	0	36.5	
3	-6	-1	-1	-43.7	
3	-4	0	-2	-20.3	..
3	-2	1	-3	3.1	--
[illegible]	[illegible]	2	-4	26.5	..
[illegible]	8	-4	0	48.7	
4	-4	-2	-1	-31.6	
4	-2	-1	-2	-9.2	--
4	0	0	-3	15.3	..
4	2	1	-4	38.7	

The unfinished tables will consider various ranges of the elements, ... consider isotopic nuclides, and will be sorted by e as well as ...



2 μL AMINO ACIDS : GEHRKE'S DERIVATIVES

100°C - 270 at 4°C/min

URINE extract + GVP - TH

R = 667

FIGURE 2

```
FILE DPH01594   USING REF FILE MPR01564
SOURCE IS urine-fraction:aminoacids(Gehrke technique
CMP ID IS GVP-17H
UNDEFINED     MASS OF MCL ION  DATE RUN 730313
RUN ON MAT 711 EXP DOWN SCAN
THRESHOLD= 4, PEAKS REJECTED   107 FOR AREA,     0 FOR WIDTH
SAMP RATE=10800 , MIN WIDTH= 2, MIN AREA= 18
(15.6 SECS, 1.49 DECS), TDEC=10.5
NUMBER OF PEAKS FOUND=170   ARRAY SPACE USED( 1750/ 8000)= 21.9%
 38 CALIB MASSES WITH LAST ONE=455.0,  6 MASSES ABOVE NOT FOUND
MISSED CALIBRATION MASSES:
281.0
******************  TABLE OF PROJECTION ERRORS **********************

 28.0      9.5   31.0     43.6   44.0      2.8   51.0     24.9   69.0    -0.3
 81.0      0.3   93.0      0.3  100.0      0.2  112.0     -0.5  119.0    -0.2
131.0     -0.2  143.0      0.1  155.0      0.1  162.0      0.1  169.0    -0.9
181.0      2.4  193.0     -2.3  205.0      1.5  212.0     -0.2  219.0    -0.9
231.0      4.1  243.0     -4.4  255.0      1.8  269.0     -0.9  293.0     1.5
305.0      0.4  319.0     -2.0  331.0      2.6  343.0     -2.2  355.0     2.6
367.0      0.6  381.0      1.1  393.0      5.1  405.0      4.4  417.0    -1.7
431.0      2.0  443.0     -4.0  455.0      6.6
MASS FILE MPH01594 HAS BEEN CREATED WITH 84 MASSES FOUND
SAMPLE BASE PEAK IS   3.7 VOLTS AND REF BASE PEAK IS 174.5 VOLTS
MATCHING TOLERANCE= 4.000 MMU FROM MASS  40 TO 700
MATCHING C 15 H 20 O  5 N  4 FL   6
   MASS        AREA     MMU ERROR     COMPOSITION
 41.03931      58.0       0.183       C  3 H  5
 43.05388       5.4      -0.899       C  3 H  7
 55.01820       3.5       1.153       C  1 H  1 N  3
                         -0.189       C  3 H  3 O  1
 55.05510      12.0       0.321       C  4 H  7
 56.06277      33.9       0.180       C  4 H  8
 57.07068      72.8       0.248       C  4 H  9
 64.01433       4.0      -1.721       C  1 H  4 O  3
                          1.863       C  2 H  2 FL   2
 68.96245       3.7
 68.97505       4.7
 69.03101       5.4      -1.698       C  2 H  3 N  3
                         -3.041       C  4 H  5 O  1
 70.03085       4.4       2.907       C  1 H  2 N  4
                          1.564       C  3 H  4 O  1 N  1
                         -3.405       H  4 N  2 FL   2
 75.00519       6.2      -1.682       H  1 O  2 N  3
                         -3.025       C  2 H  3 O  3
                          0.560       C  3 H  1 FL   2
                         -0.585       H  2 O  1 FL   3
 83.01006       5.8      -1.911       C  2 H  1 O  1 N  3
                         -3.253       C  4 H  3 O  2
                         -0.814       C  2 H  2 FL   3
 84.04382       6.0       0.230       C  2 H  4 N  4
                         -1.113       C  4 H  6 O  1 N  1
 85.02852      35.3       0.914       C  2 H  3 O  1 N  3
                         -0.429       C  4 H  5 O  2
```

		2.011	C 2 H 4 FL 3
96.00504	9.5	-0.833	C 1 H 4 C 5
		-2.170	C 2 O 1 N 4
		-3.513	C 4 H 2 O 2 N 1
		3.896	C 5 H 1 O 1 FL 1
		2.751	C 2 H 2 O 2 FL 2
		-1.073	C 2 H 1 N 1 FL 3
134.96208	3.3		
137.00035	4.6	1.639	C 5 H 1 O 3 N 2
		-2.383	C 10 H 1 O 1
		0.494	C 2 H 2 O 4 N 2 FL 1
		-2.185	C 5 O 1 N 3 FL 1
		-3.528	C 7 H 2 O 2 FL 1
		-3.330	C 2 H 1 O 2 N 3 FL 2
		2.934	H 1 O 2 N 2 FL 4
		-1.088	C 5 H 1 FL 4
		-0.890	N 3 FL 5
		-2.233	C 2 H 2 O 1 FL 5
139.02548	5.6	2.567	C 1 H 5 O 5 N 3
		-0.113	C 4 H 3 O 2 N 4
		-1.455	C 6 H 5 O 3 N 1
		-1.258	C 1 H 4 O 3 N 4 FL 1
		-2.600	C 3 H 6 O 4 N 1 FL 1
		2.129	C 7 H 3 N 1 FL 2
		2.327	C 2 H 2 N 4 FL 3
		0.984	C 4 H 4 O 1 N 1 FL 3
		-0.160	C 1 H 5 O 2 N 1 FL 4
		-3.985	C 1 H 4 N 2 FL 5
140.03549	3.2	2.065	C 4 H 4 O 2 N 4
		0.722	C 6 H 6 O 3 N 1
		-1.957	C 9 H 4 N 2
		0.920	C 1 H 5 O 3 N 4 FL 1
		-0.423	C 3 H 7 O 4 N 1 FL 1
		-3.102	C 6 H 5 O 1 N 2 FL 1
		3.162	C 4 H 5 O 1 N 1 FL 3
		-1.807	C 1 H 5 N 2 FL 5
152.03288	66.1	2.141	C 2 H 6 O 5 N 3
		-0.538	C 5 H 4 O 2 N 4
		-1.881	C 7 H 6 O 3 N 1
		-1.683	C 2 H 5 O 3 N 4 FL 1
		-3.025	C 4 H 7 O 4 N 1 FL 1
		1.704	C 8 H 4 N 1 FL 2
		1.902	C 3 H 3 N 4 FL 3
		0.559	C 5 H 5 O 1 N 1 FL 3
		0.757	H 4 O 1 N 4 FL 4
		-0.586	C 2 H 6 O 2 N 1 FL 4
153.03532	4.9	2.620	C 9 H 3 N 3
		1.278	C 11 H 5 O 1
		-3.254	C 2 H 7 O 5 N 3
		1.475	C 6 H 4 O 1 N 3 FL 1
		0.133	C 8 H 6 O 2 FL 1
		0.331	C 3 H 5 O 2 N 3 FL 2
		-3.691	C 8 H 5 N 1 FL 2
		-1.012	C 5 H 7 O 3 FL 2

		-3.494	C 3 H 4 N 4 FL 3
		2.573	C 6 H 5 FL 4
164.00092	3.5	3.877	C 5 O 3 N 4
		2.534	C 7 H 2 O 4 N 1
		-0.145	C 10 O 1 N 2
		2.732	C 2 H 1 O 4 N 4 FL 1
		1.389	C 4 H 3 O 5 N 1 FL 1
		-1.290	C 7 H 1 O 2 N 2 FL 1
		-2.435	C 4 H 2 O 3 N 2 FL 2
		-3.580	C 1 H 3 O 4 N 2 FL 3
		3.829	C 2 H 2 O 3 N 1 FL 4
		1.150	C 5 N 2 FL 4
173.98967	6.3	1.963	C 5 O 3 N 2 FL 2
		-2.059	C 10 O 1 FL 2
		0.818	C 2 H 1 O 4 N 2 FL 3
		-3.204	C 7 H 1 O 2 FL 3
		-3.006	C 2 O 2 N 3 FL 4
		3.258	C 2 N 2 FL 6
		-0.764	C 5 FL 6
174.99890	11.4	-2.896	C 7 H 1 O 3 N 3
		0.491	C 13 FL 1
		3.368	C 5 H 1 O 3 N 2 FL 2
		-0.654	C 10 H 1 O 1 FL 2
		2.223	C 2 H 2 O 4 N 2 FL 3
		-0.456	C 5 O 1 N 3 FL 3
		-1.799	C 7 H 2 O 2 FL 3
		-1.601	C 2 H 1 O 2 N 3 FL 4
		-2.943	C 4 H 3 O 3 FL 4
		0.641	C 5 H 1 FL 6
178.98990	3.5	0.599	C 7 O 3 N 2 FL 1
		-3.423	C 12 O 1 FL 1
		-0.546	C 4 H 1 O 4 N 2 FL 2
		-1.691	C 1 H 2 O 5 N 2 FL 3
		1.894	C 2 O 2 N 2 FL 5
		-2.128	C 7 FL 5
		-3.273	C 4 H 1 O 1 FL 6
180.02676	87.2	-1.582	C 6 H 4 O 3 N 4
		-2.925	C 8 H 6 O 4 N 1
		1.805	C 12 H 3 N 1 FL 1
		-2.727	C 3 H 5 O 4 N 4 FL 1
		3.339	C 6 H 6 O 4 FL 2
		2.003	C 7 H 2 N 4 FL 2
		0.660	C 9 H 4 O 1 N 1 FL 2
		-3.872	H 6 O 5 N 4 FL 2
		3.537	C 1 H 5 O 4 N 3 FL 3
		2.194	C 3 H 7 O 5 FL 3
181.03192	12.5	2.973	C 12 H 5 O 2
		3.171	C 7 H 4 O 2 N 3 FL 1
		-0.851	C 12 H 4 N 1 FL 1
		1.828	C 9 H 6 O 3 FL 1
		2.026	C 4 H 5 O 3 N 3 FL 2
		0.683	C 6 H 7 O 4 FL 2
		-0.653	C 7 H 3 N 4 FL 2
		-1.996	C 9 H 5 O 1 N 1 FL 2

		0.881	C 1 H 6 O 4 N 3 FL 3
		-0.462	C 3 H 8 O 5 FL 3
190.99983	7.6	3.119	C 7 H 1 O 4 N 3
		1.776	C 9 H 3 O 5
		-0.903	C 12 H 1 O 2 N 1
		1.974	C 4 H 2 O 5 N 3 FL 1
		-0.705	C 7 O 2 N 4 FL 1
		-2.048	C 9 H 2 O 3 N 1 FL 1
		-1.850	C 4 H 1 O 3 N 4 FL 2
		-3.193	C 6 H 3 O 4 N 1 FL 2
		1.537	C 10 N 1 FL 3
		-2.995	C 1 H 2 O 4 N 4 FL 3
196.95901	3.2		
198.03705	100.0	2.670	C 15 H 4 N 1
		-1.862	C 6 H 6 O 4 N 4
		-3.204	C 8 H 8 O 5 N 1
		2.868	C 10 H 3 N 4 FL 1
		1.525	C 12 H 5 O 1 N 1 FL 1
		-3.006	C 3 H 7 O 5 N 4 FL 1
		3.060	C 6 H 8 O 5 FL 2
		1.723	C 7 H 4 O 1 N 4 FL 2
		0.380	C 9 H 6 O 2 N 1 FL 2
		3.258	C 1 H 7 O 5 N 3 FL 3
199.03798	3.7	-0.190	C 10 H 5 O 2 N 3
		2.489	C 7 H 7 O 5 N 2
		-1.533	C 12 H 7 O 3
		-1.335	C 7 H 6 O 3 N 3 FL 1
		-2.678	C 9 H 8 O 4 FL 1
		2.052	C 13 H 5 FL 2
		-2.480	C 4 H 7 O 4 N 3 FL 2
		-3.823	C 6 H 9 O 5 FL 2
		2.250	C 8 H 4 N 3 FL 3
		0.907	C 10 H 6 O 1 FL 3
199.98764	7.7	1.838	C 9 O 4 N 2
		-2.184	C 14 O 2
		0.693	C 6 H 1 O 5 N 2 FL 1
		-3.329	C 11 H 1 O 3 FL 1
		-3.131	C 6 O 3 N 3 FL 2
		3.133	C 4 O 3 N 2 FL 4
		-0.889	C 9 O 1 FL 4
		1.988	C 1 H 1 O 4 N 2 FL 5
		-2.034	C 6 H 1 O 2 FL 5
		-1.836	C 1 O 2 N 3 FL 6
200.99631	26.7	2.681	C 9 H 1 O 4 N 2
		-1.341	C 14 H 1 O 2
		-1.143	C 9 O 2 N 3 FL 1
		1.536	C 6 H 2 O 5 N 2 FL 1
		-2.486	C 11 H 2 O 3 FL 1
		-2.288	C 6 H 1 O 3 N 3 FL 2
		-3.631	C 8 H 3 O 4 FL 2
		1.099	C 12 FL 3
		-3.433	C 3 H 2 O 4 N 3 FL 3
		3.976	C 4 H 1 O 3 N 2 FL 4
217.99628	4.4	-2.778	C 12 O 2 N 3

		-0.099	C 9 H 2 C 5 N 2
		-3.923	C 9 H 1 C 3 N 3 FL 1
		3.486	C 10 C 2 N 2 FL 2
		-0.536	C 15 FL 2
		2.341	C 7 H 1 C 3 N 2 FL 3
		-1.681	C 12 H 1 C 1 FL 3
		1.196	C 4 H 2 C 4 N 2 FL 4
		-1.483	C 7 C 1 N 3 FL 4
		-2.826	C 9 H 2 C 2 FL 4
219.99133	7.9	0.443	C 12 C 3 N 2
		-0.702	C 9 H 1 C 4 N 2 FL 1
		-1.846	C 6 H 2 C 5 N 2 FL 2
		1.738	C 7 C 2 N 2 FL 4
		-2.284	C 12 FL 4
		0.593	C 4 H 1 C 3 N 2 FL 5
		-3.429	C 9 H 1 C 1 FL 5
		-0.551	C 1 H 2 C 4 N 2 FL 6
		-3.231	C 4 C 1 N 3 FL 6
223.98842	10.9	2.617	C 11 C 4 N 2
		1.472	C 8 H 1 C 5 N 2 FL 1
		-2.550	C 13 H 1 C 3 FL 1
		-2.352	C 8 C 3 N 3 FL 2
		-3.695	C 10 H 2 C 4 FL 2
		-3.497	C 5 H 1 C 4 N 3 FL 3
		3.912	C 6 C 3 N 2 FL 4
		-0.110	C 11 C 1 FL 4
		2.767	C 3 H 1 C 4 N 2 FL 5
		-1.255	C 8 H 1 C 2 FL 5
225.02393	6.2	3.802	C 14 H 1 N 4
		-2.072	C 7 H 5 C 5 N 4
		3.994	C 10 H 6 C 5 FL 1
		2.657	C 11 H 2 C 1 N 4 FL 1
		1.315	C 13 H 4 C 2 N 1 FL 1
		1.512	C 8 H 3 C 2 N 4 FL 2
		0.170	C 10 H 5 C 3 N 1 FL 2
		-2.510	C 13 H 3 N 2 FL 2
		0.368	C 5 H 4 C 3 N 4 FL 3
		-0.975	C 7 H 6 C 4 N 1 FL 3
226.03322	10.2	-0.603	C 7 H 6 C 5 N 4
		2.784	C 13 H 5 C 2 N 1 FL 1
		2.982	C 8 H 4 C 2 N 4 FL 2
		1.639	C 10 H 6 C 3 N 1 FL 2
		-1.040	C 13 H 4 N 2 FL 2
		1.837	C 5 H 5 C 3 N 4 FL 3
		0.494	C 7 H 7 C 4 N 1 FL 3
		-2.185	C 10 H 5 C 1 N 2 FL 3
		0.692	C 2 H 6 C 4 N 4 FL 4
		-0.651	C 4 H 8 C 5 N 1 FL 4
236.99301	3.3	-0.625	C 12 H 1 C 4 N 2
		-1.770	C 9 H 2 C 5 N 2 FL 1
		1.815	C 10 C 2 N 2 FL 3
		-2.207	C 15 FL 3
		0.670	C 7 H 1 C 3 N 2 FL 4
		-3.352	C 12 H 1 C 1 FL 4

		-0.475	C 4 H 2 O 4 N 2 FL 5
		-3.154	C 7 O 1 N 3 FL 5
		-1.620	C 1 H 3 O 5 N 2 FL 6
243.99310	5.8	2.209	C 14 O 3 N 2
		1.064	C 11 H 1 O 4 N 2 FL 1
		-2.760	C 11 O 2 N 3 FL 2
		-0.081	C 8 H 2 O 5 N 2 FL 2
		-3.905	C 8 H 1 O 3 N 3 FL 3
		3.504	C 9 O 2 N 2 FL 4
		-0.518	C 14 FL 4
		2.359	C 6 H 1 O 3 N 2 FL 5
		-1.663	C 11 H 1 O 1 FL 5
		1.214	C 3 H 2 O 4 N 2 FL 6
245.00569	5.3	-1.586	C 10 H 3 O 5 N 3
		3.144	C 14 O 1 N 3 FL 1
		1.999	C 11 H 1 O 2 N 3 FL 2
		0.656	C 13 H 3 O 3 FL 2
		0.854	C 8 H 2 O 3 N 3 FL 3
		-0.489	C 10 H 4 O 4 FL 3
		-1.825	C 11 N 4 FL 3
		-3.168	C 13 H 2 O 1 N 1 FL 3
		-0.291	C 5 H 3 O 4 N 3 FL 4
		-1.634	C 7 H 5 O 5 FL 4
247.98857	4.2	2.766	C 13 O 4 N 2
		1.621	C 10 H 1 O 5 N 2 FL 1
		-2.401	C 15 H 1 O 3 FL 1
		-2.203	C 10 O 3 N 3 FL 2
		-3.546	C 12 H 2 O 4 FL 2
		-3.348	C 7 H 1 O 4 N 3 FL 3
		0.039	C 13 O 1 FL 4
		2.916	C 5 H 1 O 4 N 2 FL 5
		-1.106	C 10 H 1 O 2 FL 5
		-0.908	C 5 O 2 N 3 FL 6
248.99155	4.7	-2.087	C 13 H 1 O 4 N 2
		-3.232	C 10 H 2 O 5 N 2 FL 1
		3.032	C 8 H 2 O 5 N 1 FL 3
		0.353	C 11 O 2 N 2 FL 3
		3.230	C 3 H 1 O 5 N 4 FL 4
		-0.792	C 8 H 1 O 3 N 2 FL 4
		-1.937	C 5 H 2 O 4 N 2 FL 5
		-3.082	C 2 H 3 O 5 N 2 FL 6
254.10146	39.5	-0.052	C 10 H 14 O 4 N 4
		-1.395	C 12 H 16 O 5 N 1
		-1.197	C 7 H 15 O 5 N 4 FL 1
		3.533	C 11 H 12 O 1 N 4 FL 2
		2.190	C 13 H 14 O 2 N 1 FL 2
		2.388	C 8 H 13 O 2 N 4 FL 3
		1.045	C 10 H 15 O 3 N 1 FL 3
		-1.634	C 13 H 13 N 2 FL 3
		1.243	C 5 H 14 O 3 N 4 FL 4
		-0.100	C 7 H 16 O 4 N 1 FL 4
255.10947	3.3	0.143	C 10 H 15 O 4 N 4
		-1.200	C 12 H 17 O 5 N 1
		-3.879	C 15 H 15 O 2 N 2

M - CO_2

		-1.002	C 7 H 16 O 5 N 4 FL 1
		3.728	C 11 H 13 O 1 N 4 FL 2
		2.385	C 13 H 15 O 2 N 1 FL 2
		2.583	C 8 H 14 O 2 N 4 FL 3
		1.240	C 10 H 16 O 3 N 1 FL 3
		-1.439	C 13 H 14 N 2 FL 3
		1.438	C 5 H 15 O 3 N 4 FL 4
255.99071	4.4	3.833	C 10 O 5 N 4
		-0.189	C 15 O 3 N 2
		-1.334	C 12 H 1 O 4 N 2 FL 1
		-2.479	C 9 H 2 O 5 N 2 FL 2
		3.785	C 7 H 2 O 5 N 1 FL 4
		1.106	C 10 O 2 N 2 FL 4
		-2.916	C 15 FL 4
		3.983	C 2 H 1 O 5 N 4 FL 5
		-0.039	C 7 H 1 O 3 N 2 FL 5
		-1.184	C 4 H 2 O 4 N 2 FL 6
256.94409	19.5		
261.98169	7.4	-0.900	C 11 O 4 N 2 FL 2
		-2.045	C 8 H 1 O 5 N 2 FL 3
		0.395	C 6 O 3 N 2 FL 6
		-3.627	C 11 O 1 FL 6
262.99121	19.3	2.023	C 14 O 3 N 2 FL 1
		0.878	C 11 H 1 O 4 N 2 FL 2
		-2.946	C 11 O 2 N 3 FL 3
		-0.267	C 8 H 2 O 5 N 2 FL 3
		3.318	C 9 O 2 N 2 FL 5
		-0.704	C 14 FL 5
		2.173	C 6 H 1 O 3 N 2 FL 6
		-1.849	C 11 H 1 O 1 FL 6
269.98486	4.2	1.101	C 13 H 1 O 5 N 1 FL 1
		1.299	C 8 O 5 N 4 FL 2
		-2.723	C 13 O 3 N 2 FL 2
		-3.868	C 10 H 1 O 4 N 2 FL 3
		3.541	C 11 O 3 N 1 FL 4
		2.396	C 8 H 1 O 4 N 1 FL 5
		2.594	C 3 O 4 N 4 FL 6
		1.251	C 5 H 2 O 5 N 1 FL 6
		-1.428	C 8 O 2 N 2 FL 6
273.98608	10.7	-2.749	C 14 O 4 N 3
		-3.894	C 11 H 1 O 5 N 3 FL 1
		3.515	C 12 O 4 N 2 FL 2
		2.370	C 9 H 1 O 5 N 2 FL 3
		-1.652	C 14 H 1 O 3 FL 3
		-1.454	C 9 O 3 N 3 FL 4
		-2.797	C 11 H 2 O 4 FL 4
		-2.599	C 6 H 1 O 4 N 3 FL 5
		-3.942	C 8 H 3 O 5 FL 5
		0.788	C 12 O 1 FL 6
********** MISSED CALIBRATION MASS=281.0 **********			
281.08472	4.7	-3.910	C 11 H 13 O 5 N 4
		3.499	C 12 H 12 O 4 N 3 FL 1
		2.157	C 14 H 14 O 5 FL 1
		0.820	C 15 H 10 O 1 N 4 FL 1

		2.354	C 9 H 13 O 5 N 3 FL 2
		-0.325	C 12 H 11 O 2 N 4 FL 2
		-1.668	C 14 H 13 O 3 N 1 FL 2
		-1.470	C 9 H 12 O 3 N 4 FL 3
		-2.812	C 11 H 14 O 4 N 1 FL 3
		1.917	C 15 H 11 N 1 FL 4
282.09814	11.1	1.728	C 11 H 14 O 5 N 4
		-3.439	C 13 H 15 O 4 N 2 FL 1
		3.970	C 14 H 14 O 3 N 1 FL 2
		2.825	C 11 H 15 O 4 N 1 FL 3 M-OB
		0.146	C 14 H 13 O 1 N 2 FL 3
		3.023	C 6 H 14 O 4 N 4 FL 4
		1.681	C 8 H 16 O 5 N 1 FL 4
		-0.999	C 11 H 14 O 2 N 2 FL 4
		-2.144	C 8 H 15 O 3 N 2 FL 5
282.99707	5.3	1.422	C 14 H 1 O 3 N 2 FL 2
		0.277	C 11 H 2 O 4 N 2 FL 3
		-2.403	C 14 O 1 N 3 FL 3
		-3.547	C 11 H 1 O 2 N 3 FL 4
		-0.868	C 8 H 3 O 5 N 2 FL 4
		3.861	C 12 O 1 N 2 FL 5
		2.717	C 9 H 1 O 2 N 2 FL 6
		-1.305	C 14 H 1 FL 6
285.98682	9.3	-2.073	C 15 O 4 N 3
		-3.218	C 12 H 1 O 5 N 3 FL 1
		3.046	C 10 H 1 O 5 N 2 FL 3
		-0.976	C 15 H 1 O 3 FL 3
		-0.778	C 10 O 3 N 3 FL 4
		-2.121	C 12 H 2 O 4 FL 4
		-1.923	C 7 H 1 O 4 N 3 FL 5
		-3.266	C 9 H 3 O 5 FL 5
		1.464	C 13 O 1 FL 6
		-3.068	C 4 H 2 O 5 N 3 FL 6
290.98560	4.6	1.376	C 15 O 4 N 2 FL 1
		0.231	C 12 H 1 O 5 N 2 FL 2
		-3.593	C 12 O 3 N 3 FL 3
		2.671	C 10 O 3 N 2 FL 5
		-1.351	C 15 O 1 FL 5
		1.526	C 7 H 1 O 4 N 2 FL 6
		-2.496	C 12 H 1 O 2 FL 6
293.98828	4.7	0.548	C 15 O 3 N 2 FL 2
		-0.597	C 12 H 1 O 4 N 2 FL 3
		-1.742	C 9 H 2 O 5 N 2 FL 4
		1.843	C 10 O 2 N 2 FL 6
		-2.179	C 15 FL 6
297.97510	3.2	3.794	C 15 O 5 FL 2
		-1.175	C 12 O 4 N 1 FL 4
		-2.320	C 9 H 1 O 5 N 1 FL 5
		-2.122	C 4 O 5 N 4 FL 6
311.97388	4.9	-0.576	C 15 O 5 N 1 FL 2
		0.719	C 10 O 4 N 1 FL 6
312.98804	11.1	1.882	C 15 O 3 N 2 FL 3
		0.737	C 12 H 1 O 4 N 2 FL 4
		-3.087	C 12 O 2 N 3 FL 5

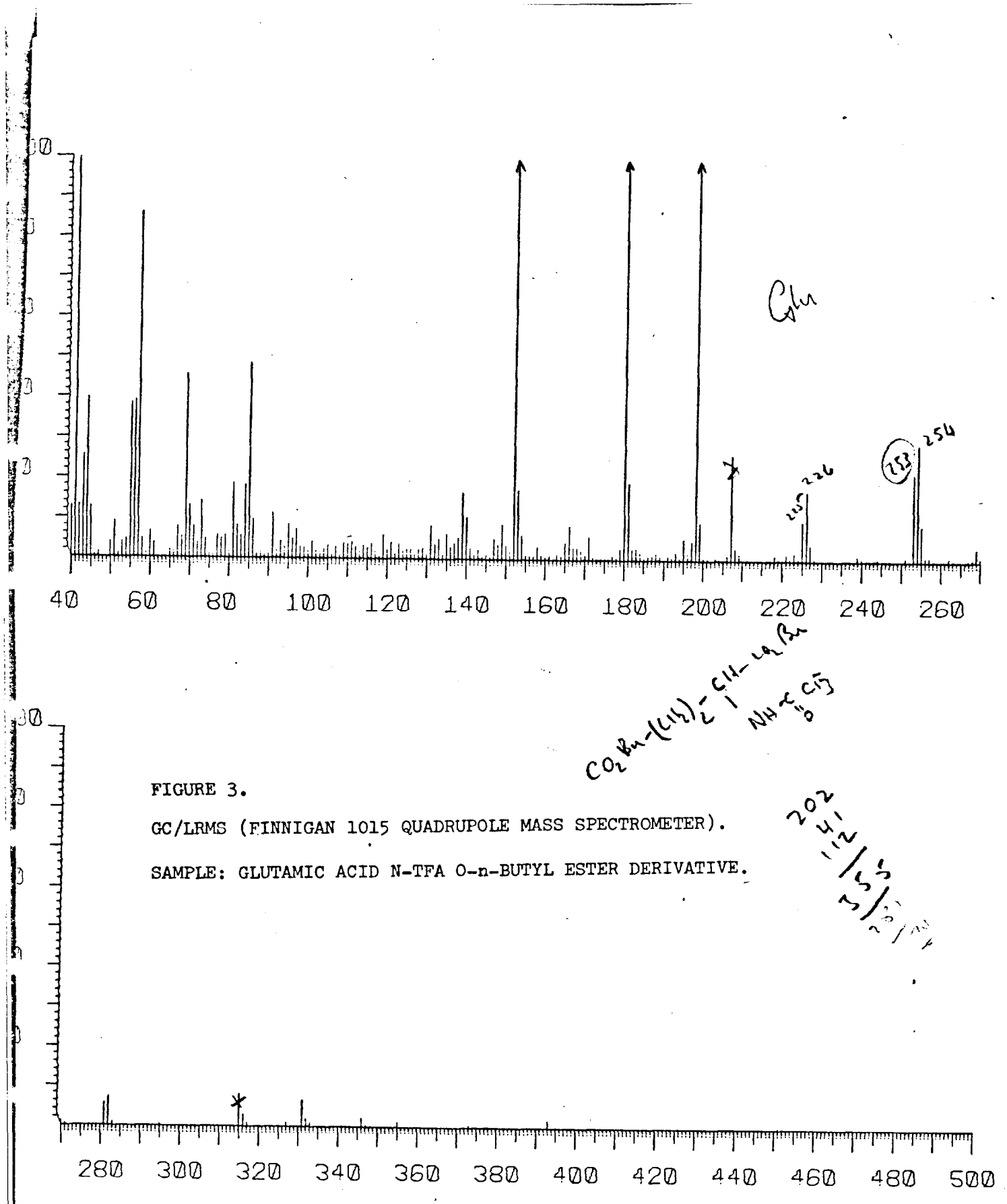


FIGURE 3.

GC/LRMS (FINNIGAN 1015 QUADRUPOLE MASS SPECTROMETER).

SAMPLE: GLUTAMIC ACID N-TFA O-n-BUTYL ESTER DERIVATIVE.

APPENDIX B

LETTERS OF INTEREST

STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

DEPARTMENT OF CHEMISTRY

December 17, 1973

Professor Carl Djerassi
Department of Chemistry
Stanford University
Stanford, California 94305

Dear Carl:

I am writing to indicate the anticipated use of mass spectral facilities by my research group in the foreseeable future. As has been true in the past, we plan to utilize both GC/HRMS and simple HRMS for various purposes, especially 1) the determination of structure of enzymic cyclization products, including members of the lanosterol class, derived from squalene oxide-like substrates, the purpose being the elucidation of the mechanism of enzymic steroid synthesis, and 2) the characterization and confirmation of structures of intermediates in the synthesis of natural products, including polycyclic terpenoids, alkaloids of physiological interest, and nucleosides,and 3) identification and/or structure determination of organic materials employed in our organic-inorganic program devoted to nitrogen fixation and related processes.

Very truly yours,

EEvT/jlb

121

DATE: December 3, 73

TO : Carl Djerassi

FROM : Keith Hodgson

SUBJECT: Response to inquiry about GC/HRMS facility

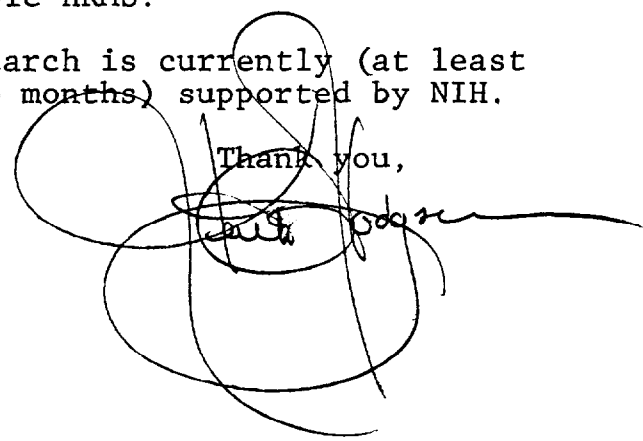
In response to your three questions concerning the potential use of upgraded GC/HRMS facilities:

1. Yes, especially in the study of certain biological ligands and lower molecular weight ligand-metal complexes.

2. Potential use of the facility might run in the range of 8-10 samples per year most of which probably would be handled most easily by simple HRMS.

3. No, no research is currently (at least for the next 6 months) supported by NIH.

Thank you,



STANFORD UNIVERSITY • OFFICE MEMORANDUM • STANFORD UNIVERSITY • OFFICE MEMORANDUM

DATE: December 3, 1973

TO : Professor Carl Djerassi

FROM : James P. Collman
Professor of Chemistry

SUBJECT:

Please excuse our belated response to your inquiry of November 20 concerning a potential upgrading of mass spectrometry facilities. The service you mention in your memo of the 20th would be valuable to us. We would have significant use for the GC/HRMS for a project dealing with models for cytochrome P_{450} based monooxygenases currently supported by the NIH.

JPC:lb

James P. Collman

DATE: December 13, 1973

To : Professor Djerassi

FROM : Professor Harry S. Mosher

SUBJECT: Your proposal to the NIH

On our NIH Grant on the investigation of animal toxins we have been studying natural products from the skin of Central American frogs (atelopidtoxin) and some products from marine animals (nudibranchs) as well as some new choline esters isolated from the hypobranchial gland of various sea snails. Some, if not all, of these are mixtures. Obviously the new capabilities of the mass spectrometry laboratory would be of value to me. I expect only occasional use of HRMS and GC/HRMS, but on these occasions these techniques would be very important.

Harry S. Mosher

• STANFORD UNIVERSITY • OFFICE MEMORANDUM • STANFORD UNIVERSITY • OFFICE MEMORANDUM • STANFORD UNIVERSITY • OFFICE MEMORANDUM

DATE: 14 December 1973

TO : C. Djerassi

FROM : W. S. Johnson

SUBJECT:

The contemplated new facility for high resolution mass spectrometry and combined gas chromatography/high resolution mass spectrometry would be of extreme value to our research program concerning the non-enzymic biogenetic-like cyclization of polyolefines, a project which is presently supported by NIH Grant AM 3787-14. If this facility were to become available, we would expect to use it extensively in the analysis of product mixtures of the aforementioned cyclizations. We estimate that our need for the gas chromatographic capability would be about 20% of the total need for the mass spectrometry service.

W. S. Johnson cs

STANFORD UNIVERSITY MEDICAL CENTER
STANFORD, CALIFORNIA 94305 • (415) 321-1200 EXT. 5785

STANFORD UNIVERSITY SCHOOL OF MEDICINE
Department of Anesthesia

November 30, 1973

Professor Joshua Lederberg
Department of Genetics
School of Medicine
Stanford University
Stanford, California 94305

Dear Dr. Lederberg:

Thank you for including my laboratories in the group which could be served by a GC/HRMS facility. As you know, Dr. Cohen and I have our own GC/MS/ Computer System. Our use of the proposed facility would be limited to those times when it is necessary to use high resolution ot identify a metabolite. I would estimate a need for three GC/HRMS and three HRMS Spectra per year. My work is entirely supported by the National Institutes of Health.

Sincerely yours,
James R. Trudell by RW
James R. Trudell, Ph.D.

JRT:rw

126

NOV 27 1973

VETERANS ADMINISTRATION
HOSPITAL
3801 MIRANDA AVENUE
PALO ALTO, CALIFORNIA 94304

IN REPLY
REFER TO:

November 26, 1973

Professor J. Lederberg
Department of Genetics
Stanford University School of
Medicine
Stanford, California 94305

Dear Prof. Lederberg:

Dr. Allan Duffield of your department has informed me that you plan to obtain additional apparatus that would provide high resolution GC/MS as a service to the Stanford community.

We have in the past used the hospitality of your department in the identification of metabolites and derivatives of phenothiazine drugs and cannabinoids by GC/MS. Originally, we had the collaboration of Dr. B. Halpern and more recently Dr. A. Duffield, who was instrumental in helping us with some of our problems.

Our department would indeed be most interested in availing ourselves of GC/MS analyses in the course of our current NIH projects which again are concerned essentially with drug metabolism and the isolation and characterization of unknown drug derivatives.

As a rough estimate, I would think that we may be interested in the analyses of about five samples per month, two of which will require high resolution MS.

I certainly hope that your project to acquire the sophisticated new instrumentation you are seeking will be successful.

Sincerely yours,

I. S. Forrest

Irene S. Forrest, Ph.D.
Chief, Biochem. Research Lab.
(151F)

ISF:jr

Show veteran's full name, VA file number, and social security number on all correspondence.

127

DEC 3 1973

DATE: November 30, 1973

TO : Joshua Lederberg

FROM : I. Rabinowitz, Ph.D.
D.I. Wilkinson, Ph.D.

SUBJECT: RE: NIH GC/HRMS Proposal

Research carried out in this department has strongly implicated a role for the prostaglandins in the etiology of psoriasis (E. M. Farber, K. Aso, 32nd Annual Meeting, American Academy Dermatology, Chicago, Ill., Dec. 1973; E. M. Farber et al, J. Invest. Derm., in preparation; E. M. Farber et al, Nature New Biology, in preparation). The prostaglandins are a class of C_{20} fatty acids, having molecular weights near 350 and basal tissue concentrations in the nanogram and picogram per gram range. The prostaglandins are presently detected by radioimmunoassay, bioassay and mass spectrometric techniques, among others. There is considerable controversy concerning the method of choice for measurement of absolute amounts of prostaglandin in various tissues. In particular, it has been suggested that mass spectrometric techniques yield more accurate quantitative assays than radioimmunoassay techniques (Adv. Biosciences, 9, 71-123, 1973, Ed. G. Raspé, S. Bernhard, Pergamon Press, N.Y.). Radioimmunoassay techniques are currently in use in our laboratories, and the addition of mass spectrometry capability would greatly increase the definitiveness of our studies, as well as make available to us a powerful tool for the study of prostaglandin precursors and metabolites. Work to date has been supported in part by NIH Grant No. AM 15107.

Israel Rabinowitz
I. Rabinowitz, Ph.D.
Department of Dermatology

David I. Wilkinson
D. I. Wilkinson, Ph.D.
Department of Dermatology

IR:DIW:ss

DATE: November 30, 1973

TO : Joshua Lederberg, Department of Genetics
Carl Djerassi, Department of Chemistry

FROM : Eugene D. Robin, M.D., Department of Respiratory Medicine

SUBJECT: Your memo of November 20, 1973 describing a proposed GC/HRMS facility.

I have applied to the NIH for a continuation of my research grant, Adaptations To 0_2 Depletion in which I have proposed to measure the redox state of NAD^+/NADH and $NADP^+$/NADPH by measuring the ratio of oxidized to reduced redox pairs using g gas chromatography/mass spectrometry. These analyses will be conducted with the assistance of Drs. Alan Duffield and Wilfred Pereira of the Department of Genetics. I welcome the opportunity to have a GC/HRMS facility available on campus to support the GC/LRMS available in the department of genetics. The facility you propose to establish will be of importance to us in those instances where assignment of molecular composition to ionized fragments is crucial for mass spectral interpretation. I would anticipate using this service between one and two times a month.

Sincerely yours,

Gene

Eugene D. Robin, M.D.
Professor of Medicine and Physiology

EDR:ods

NOV 30 1973

VETERANS ADMINISTRATION
HOSPITAL
3801 MIRANDA AVENUE
PALO ALTO, CALIFORNIA 94304

November 28, 1973

IN REPLY
REFER TO:

Dr. Joshua Lederberg
Department of Genetics
Stanford University School of Medicine
Palo Alto, California 94305

Dear Dr. Lederberg:

I should be very pleased if you were able to obtain through the National Institutes of Health a GC/MS facility which could be shared jointly by members of the Stanford University faculty.

At present, I am being funded under grant DA-00424-01 for a study of the metabolism of marihuana. We have made significant progress in our methods of extracting metabolites, in isolating new ones by thin-layer chromatographic techniques, and by purifying them to some degree as determined by GLC. The big bottleneck has been the lack of ready access to a GC/MS set-up which would permit further characterization of the metabolites.

Our needs would be primarily for GC/low resolution MS, for which we have extensive need, perhaps the analysis of 15-25 samples per month. Depending on the outcome of these analyses, we might have 1 to 2 samples per month requiring GC/high resolution MS. We anticipate having little need for high resolution MS without GC because of the fact that our samples are isolated from complex mixtures and are nearly impossible to purify.

If there is any way in which I could assist in helping obtain such a facility for the University, please let me know.

Sincerely yours,

Leo E. Hollister, M.D.
Associate Professor of Medicine

LEH:bh

Show veteran's full name, VA file number, and social security number on all correspondence.

130

STANFORD UNIVERSITY HOSPITAL
Pharmacy Department

Date September 5, 1973

To: Dr. J. Lederberg, Director
Department of Genetics

From: Hiram H. Sera, Director

Subject: Drug Analysis Service with Gas Chromatograph and Mass Spectrometer.

I wish to express our appreciation to your department for assisting us in identifying a drug sample submitted to us from the E1 patient care area.

BACKGROUND:

The patient on E1A with G.I. disturbance, joint pains and occasional spike temperature was found to possess an unidentified medication in a plastic vial and was found to have self-administered the drug intramuscularly while in the hospital. The house staff was notified and the drug sample was submitted to us for immediate identification.

Through my previous association and knowledge of Drs. Summons' and W. Perieras' (in Dr. Duffield's instrumentation research laboratory) work with gas chromatograph and mass spectrometer, I had taken the liberty to request their assistance in the identification.

In an hour, the determination was made and the drug was found to be Pentazocaine or Talwin which is a synthetic analgesic used commonly in this hospital in tablet and injection forms.

Since we do occasionally receive similar requests from physicians, I wish to call on your staff again in the future. Thank you.

HHS:lh
cc: Mr. John Williams
Dr. Roger Summons
Dr. W. Periera
Dr. A. Duffield

DATE: November 26, 1973

TO : Joshua Lederberg, Department of Chemistry
Carl Djerassi, Department of Chemistry

FROM : Sumner M. Kalman

SUBJECT: Mass Spectrometry, Your Memo of November 20, 1973.

A central facility for mass spectrometry and GC/MS would be highly desirable from my point of view. We often need to identify metabolites of drugs that interfere with our assays, and that represent research problems as well. Frequently we need to check the purity of a reference material which is in short supply. I have received much help from both your laboratories in the past and would welcome the opportunity to use an expanded facility. For many of our problems low resolution MS is satisfactory and I hope you mean to provide this service too.

With respect to your questions I anticipate that

(1) Yes.

(2) We would probably use GC/MS once a month or more. We would use MS at about the same rate.

(3) Yes.

Sincerely yours,

Sumner M. Kalman M.D.

Sumner M. Kalman, M.D.
Professor of Pharmacology
Director, Drug Assay Lab

DEC 3 1973

DATE: 3 December 1973

TO : Joshua Lederberg

FROM : Jack Barchas

SUBJECT: GC/HRMS

Our thanks to you and Alan Duffield for inquiring of our interest in the proposed GC/HRMS. We would find it quite useful, as we are currently applying for funding for a quadrupole mass spectrometer for mass fragmentography studies. With such a unit, there would be many times when the capability of the HRMS instrumentation would be valuable in structural elucidation. We would expect very heavy utilization of our instrument if we were to obtain the funding, and, therefore, would expect to make considerable use of the proposed GC/HRMS, which is an essential ancillary tool.

The GC aspects of the instrument would be valuable, since we would expect to be studying a number of unknowns and the GC separation would be an integral part of that process.

Our work is supported by NIMH, ONR, NASA, and the Alcohol Abuse division of HEW.

Jack

JDB/rs

• STANFORD UNIVERSITY • OFFICE MEMORANDUM • STANFORD UNIVERSITY • OFFICE MEMORANDUM • STANFORD UNIVERSITY • OFFICE MEMORANDUM

NOV 30 1973

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION
AMES RESEARCH CENTER
MOFFETT FIELD, CALIFORNIA 94035

REPLY TO
ATTN OF: LPE: 239-9

November 28, 1973

Professor Joshua Lederberg
Department of Genetics
School of Medicine
Stanford University
Stanford, CA 94305

Dear Professor Lederberg:

I was delighted to learn of your proposed plans to upgrade your mass spectrometry capabilities by providing routine high resolution mass spectrometry (HRMS) and combined gas chromatography/high resolution mass spectrometry (GC/HRMS). Such a service could be of inestimable value to our program. As you know we are developing gas chromatography/high resolution mass spectrometry facilities for NASA's interests. In particular we are modifying our equipment in order to determine carbon and nitrogen isotopic compositions of organic molecules. If available we would use your proposed facilities for our routine GC/HRMS analysis of biologically significant molecules which are sought in our program. Most of our work requires GC/HRMS as opposed to HRMS. In addition, we are also most interested in computer programs which aid in mass spectral interpretations. Although we have a few of our own programs, we would be most eager to upgrade our own interpretation capabilities through use of programs from your facility.

Our work thus far has been supported solely by NASA; we are not supported at present by NIH.

I hope that our expression of interest will be of use to you in obtaining funding for a potentially most useful analytical facility.

Sincerely yours,

Keith A. Kvenvolden
Chief, Chemical Evolution
Branch

DATE: November 28, 1973

TO : Joshua Lederberg, Ph.D.
S331

FROM : William R. Fair, M.D.
S287

SUBJECT: Use of facilities for high resolution mass spectral analysis with gas chromatography.

As your memo of November 1973 requested, we have answered the questions concerning our interest in GC/HRMS.

1. This service would be of definite value to us in two projects currently being investigated in our laboratories. a) The identification, distribution, and biological significance of the prostatic antibacterial factor (PAF). Our preliminary experiments indicate that this is a basic polypeptide, perhaps attached to a divalent metal such as zinc. b) This service would also be of value in the determination of the urinary polyamine levels in patients with various genitourinary tract malignancies. Our initial experiments along this line indicate that there is significant elevation of polyamines in patients with prostatic carcinoma. The use of GC/HRMS would enable a more precise quantitation of these differences and enable us to expand our research into other areas concerning the biochemical significance of the polyamines.

2. I would estimate that on the PAF project we would use approximately 2-4 samples per month and perhaps 10-12 samples per month on the polyamine projects. Both of these projects would require the use of GC/HRMS.

3. A portion of our research on the PAF is currently supported by a grant from the NIH. The amount of this grant is $36,698, and this grant will terminate on December 31, 1974.