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: Biomolecular Synthesis

2. PRINCIPAL INVESTIGATOR

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

Wipke, W. Todd

2B. TITLE OF POSITION

Associate Professor of Chemistry

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

Natural Sciences II
University of California
Santa Cruz, California 95064

2D. DEGREE: Ph.D.

2E. SOCIAL SECURITY NO.: [redacted]

2F. TELEPHONE DATA — Area Code: 408 — TELEPHONE NUMBER AND EXTENSION: 429-2397

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

Chemistry Board of Studies

2H. MAJOR SUBDIVISION *(See Instructions)*

Division of Natural Sciences

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

FROM: Oct. 1, 1976 — THROUGH: Sept. 30, 1979

4. TOTAL DIRECT COSTS REQUESTED FOR PERIOD IN ITEM 3: $ 391,532

5. DIRECT COSTS REQUESTED FOR FIRST 12-MONTH PERIOD: $ 172,084

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

Natural Sciences II
University of California
Santa Cruz; California 95064

16th Congressional District

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

A. [X] 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

N/A

TO BE COMPLETED BY RESPONSIBLE ADMINISTRATIVE AUTHORITY *(Items 9 through 13 and 15B)*

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

The Regents of the University of California
University of California, Santa Cruz
Santa Cruz, California 95064
IRS No. 1-94-1539563-A1
16th Congressional District

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

Léo F. Laporte
Dean Division of Natural Sciences

Telephone Number(s): (408) 429-2931

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

[ ] FEDERAL [X] STATE [ ] LOCAL [ ] OTHER *(Specify)*

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

H. J. Zenner
Contracts and Grants Officer
University of California
Santa Cruz, California 95064
(408) 429-2775

Telephone Number

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

20 - Division of Natural Sciences

14. PHS ACCOUNT NUMBER *(Enter if known)*

45 - 1481

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 *(Signature required on original only. Carbon, "Per" signatures not acceptable)*	A. SIGNATURE OF PERSON NAMED IN ITEM 2A	DATE
	B. SIGNATURE OF PERSON NAMED IN ITEM 10	DATE

PHS 398

SECTION 1

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE
PUBLIC HEALTH SERVICE

RESEARCH OBJECTIVES

LEAVE BLANK

PROJECT NUMBER

NAME AND ADDRESS OF APPLICANT ORGANIZATION

The Regents of the University of California
University of California, Santa Cruz, CA 95064

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

W. Todd Wipke	[redacted]	Associate Professor	Chemistry Board of Studies
Graham M. Smith	[redacted]	Research Associate	Chemistry Board of Studies
Hartmut Braun	none	Research Associate	Chemistry Board of Studies
S. Krishnan	none	Research Associate	Chemistry Board of Studies
Glenn I. Ouchi	[redacted]	Research Assistant	Chemistry Board of Studies

TITLE OF PROJECT

RESOURCE-RELATED RESEARCH: BIOMOLECULAR SYNTHESIS

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

The objectives of this research are to develop the logical and heuristic principles of biomolecular synthesis and incorporate this information into a practical computer program to assist investigators in designing organic syntheses of complex biomolecules. Special emphasis is placed on stereospecific syntheses using the strategies involving steric effects, electronic effects, and strain energy, as well as symmetry, graph theory, and topology. The methods involve computer graphics input/output to the chemist, three-dimensional model building and analysis, strategic plan formation, selection of relevant chemical transforms, and evaluation of generated synthetic schemes. The project is an extension of the SECS Simulation and Evaluation of Chemical Synthesis program, and is proposing to collaborate with the Stanford SUMEX resource.

LEAVE BLANK

SECTION II – PRIVILEGED COMMUNICATION T. Wipke [redacted]

DETAILED BUDGET FOR FIRST 12-MONTH PERIOD

FROM: October 1, 1976 THROUGH: September 30, 197

DESCRIPTION (Itemize)		TIME OR EFFORT %/HRS.	AMOUNT REQUESTED (Omit cents)		
PERSONNEL					
NAME	TITLE OF POSITION		SALARY	FRINGE BENEFITS	TOTAL
Wipke, W. Todd, Ph.D.	PRINCIPAL INVESTIGATOR				
	2 summer months	100%	4,704	31	4,735
Braun, H., Ph.D.	Postdoc. II All year	100%	12,105	1,816	13,921
Smith, G., Ph.D.	Postdoc. I All year	100%	11,598	1,740	13,338
Unnamed, Ph.D.	Postdoc. I All year	100%	11,142	1,671	12,813
Unnamed	Programming Asst. equiva-				
	lent to one Prog. Asst. II				
	All year Approximately-	57%	6,000	40	6,040
Ouchi, Glen, M.S.	Two (2) Grad. Students				
Unnamed	(Res. Assists.) all year	50%	10,596	70	10,666
Unnamed	One secretary I All Year	50%	4,011	602	4,613
	Totals		60,156	5,970	66,126

CONSULTANT COSTS		0
EQUIPMENT	(See attached equipment list)	74,338
SUPPLIES	Magnetic tape, electrostatic paper, movie film, office supplies	3,000
TRAVEL – DOMESTIC	Trips to Conference and SUMEX site	1,500
TRAVEL – FOREIGN	One overseas trip to present paper and visit computer synthesis projects, round trip to Burgenstock, Switz.	1,500
PATIENT COSTS (See instructions)		0
ALTERATIONS AND RENOVATIONS		0
OTHER EXPENSES (Itemize)	Two leased lines to SUMEX resource (4,500); publication costs (3,000); manuals and documentation (1000); DEC maintenance of equipment (8,000); Telephone equip. rental, long distance, postage(2000); Computer time IBM 360/40 (1,000); Disk drive lease (6,120)	25,620
TOTAL DIRECT COST (Enter on Page 1, Item 5)		172,084

INDIRECT COST (See Instructions)

DATE OF DHEW AGREEMENT: March 26, 1975

_____ % S&W*

34.2 % TDC* Modified

☐ WAIVED

☐ UNDER NEGOTIATION WITH: _____

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

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

DESCRIPTION		1ST PERIOD (SAME AS DETAILED BUDGET)	ADDITIONAL YEARS SUPPORT REQUESTED *(This application only)*					
			2ND YEAR	3RD YEAR	4TH YEAR	5TH YEAR	6TH YEAR	7TH YEAR
PERSONNEL COSTS		66,126	71,574	77,874				
CONSULTANT COSTS *(Include fees, travel, etc.)*		0	0	0				
EQUIPMENT		74,338	3,000	3,000				
SUPPLIES		3,000	3,000	3,000				
TRAVEL	DOMESTIC	1,500	1,500	1,500				
	FOREIGN	1,500	1,500	1,500				
PATIENT COSTS		0	0	0				
ALTERATIONS AND RENOVATIONS		0	0	0				
OTHER EXPENSES		25,620	26,000	26,000				
TOTAL DIRECT COSTS		172,084	106,574	112,874				
TOTAL FOR ENTIRE PROPOSED PROJECT PERIOD *(Enter on Page 1, Item 4)* →						$ 391,532		

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.)*

Budget Justification

Personnel

Summer salary is requested for Professor Wipke. Also, funds for three postdoctoral fellows are requested. Drs. Braun, Smith, and Krishnan are currently working in this research group. These postdoctoral positions provide training to the students in computer synthesis and they provide a high level of chemical knowledge needed in this work. Support is requested for graduate student Glenn Ouchi, who is already working on this project, and one other graduate student yet to be named. Graduate students are important to this project because they provide needed continuity. Additional programming assistance will be provided by undergraduates, especially by juniors and seniors continuing their research work during the summer. This provides many premedical students valuable computer experience. Also requested is partial support (50%) of a secretary for typing manuscripts, documentation, and project-related correspondence with users.

Salaries were figured to include anticipated 5% range adjustment in each new fiscal year plus normal promotions or merit increases when appropriate.

EQUIPMENT LIST

GT44	Graphics terminal	$34,500
DL11-E	Asynchronous link to host computer	595
	Interface to 9 track magnetic tape drive	3,000
DD11-B	Systems Unit Cage	275
LV11BA	Electrostatic printer/plotter	12,400
GP-3-3D	Graphic Tablet and model 1454 interface	6,700
	16mm movie camera with animation motor	4,000
	2 Teletype terminals (1 Data Media, 1 Thermal Printing	5,000
	2 pair VA 3405 Vadic 1200 baud modems with VA 1601 enclosure and power supply	3,660
	Subtotal	$70,130
	California State Sales Tax -6%	4,208
	Total Cost	$74,338

Equipment

This project plans to use the SUMEX Resource at Stanford University (directed by Professor J. Lederberg, Department of Genetics) for computing, file storage and scientific collaboration. SUMEX is the Stanford University Medical Experimental Computer System which has as its objectives the application of advanced computer science to the field of medicine and the exploration of collaborative interactions between active researchers in this area which are facilitated by computer networking. The Biomolecular Synthesis project described in this proposal is an application of artificial intelligence to chemical synthesis problems in medicine and consequently it has much in common with other projects at SUMEX.

For the past 6 months we have been users of the SUMEX Resource over a leased and a dial-up line. File storage is a major problem at SUMEX. The total file space currently is 120 K pages (1 page = 512 words of 36 bits each) which is completely allocated. We have had an allocation of 4 K pages, but requested 8 K pages. To ease this problem for the entire SUMEX community, we request funds for an RP03 disk drive (20 K pages) to attach to the last available slot in the disk controller at SUMEX for a 17% increase in on-line storage. This is shown in the budget as a 3-year lease plan under category "other expenses".

The remainder of the equipment requested is to facilitate efficient use of people and SUMEX from Santa Cruz. We have found graphical input and output to be the most efficient means of interfacing chemist to computer because structural diagrams are the natural language of the chemist. Since this project involves computer construction and analysis of three-dimensional molecular models, we have a need to be able to visualize these models, to rotate them, and to move atoms in three-space by hand (to change conformation). At Princeton Dr. Wipke developed a 3-D acoustic tablet for modifying such models and had available an LDS-1 display system for rotation of the models. When he moved to Santa Cruz that equipment remained behind. Currently we are using a GT40 display which has a small screen (12") and is too slow to be able to rotate three-dimensional structures to create a kinetic depth effect.

The requested GT44 graphics terminal is 3 times faster than the GT40 and can rotate small molecules ($<$ 72 atoms) in real time. Additionally it has a larger screen (17"). Since it is software compatible with the GT40, we will be able to do two-dimensional graphics on either terminal. Currently time is lost waiting to use the GT 40. This is because we are building on an operational graphics program and more time is spent in graphical debugging and testing. The GT44 would provide more available display time. The GT44 is also the more sensible terminal to which the requested peripherals can be interfaced. The two options 1) expand the existing GT40 to accommodate the peripherals (software and hardware required) and 2) purchase GT44 and connect peripherals to GT44, are very close in cost. Option 2 provides two graphics terminals instead of one, gives us the needed 3-D capability, more flexibility in the event of hardware failure, and basically a more supportable, reliable system.

The GP-3-3D is an acoustic 3-D tablet now marketed by Science Assessories. It is used for 2-dimensional drawing as well as 3-D moving of atoms and tracing of 3-D models. It is really the only efficient way to change the conformation of a structure. We used one extensively at Princeton (the prototype) and are severely hampered now without it. It is a necessity for this work.

The DL11-E asynchronous link connects the GT44 to the Vadic modem. Two pair of Vadic modems are requested, one pair for each of two leased lines connecting the two graphics terminals to SUMEX. That is, this grant would cover all the cost of communication right up to the SUMEX computer including the modems at SUMEX. Vadic modems seem to be extremely reliable. (TYMNET is also using Vadic modems.)

The two teletype terminals are needed for routine text editing of programs. Although the market is changing rapidly, at this time a Data Media terminal seems

GPO : 1969 O - 350-360

a good choice, because it is the standard of SUMEX, meaning maintenance is easier, and SUMEX supports TV-edit which uses the capabilities of the terminal to advantage. The other terminal would be a portable thermal printing terminal that could be checked out for remote use to encourage use during the graveyard shift.

A 9-track magnetic tape drive from the University of California is available at no cost. Information Sciences here has designed and tested an interface. Funds are requested to duplicate their operational interface. This tape drive will allow us to rapidly transmit to, and receive files from SUMEX, eliminating mail delays of several days. We have a number of infrequently used programs which can be loaded only when needed. Also, we will better be able to send tapes of programs requested by users. The DD11-B Systems unit cage is needed for the tape drive interface.

The electrostatic printer/plotter is needed to provide rapid listings locally without again waiting days for the mail from Stanford to arrive. This device will be used also for graphically recording syntheses, plotting 3-D models, and preparation of documentation. It is quiet, and highly reliable with little maintenance required. The high resolution model was selected because it is needed for plotting.

The movie camera will be used to film graphical results, for documentation and recording interaction as well as producing demonstration films and teaching films. Films are even superior to on-line demonstrations for presentations to large audiences and are essential to describing an interactive-graphics program.

In years 02 and 03

$3000 is for additional equipment, for example an additional teletype or interactive device such as switches, knobs, or a color wheel for making color films of graphical displays.

Domestic travel is needed for presenting the results of this work at national meetings and for travel to the SUMEX site. The DENDRAL and SECS projects have joint group meetings monthly.

Foreign travel is requested to present the results of this work overseas at invited lectures. The Buergenstock Conference is especially important since it is concerned mainly with stereochemistry. Additionally, Dr. Francois Choplin, Strasbourg, France, is modifying SECS for inorganic chemistry and periodic visits will be helpful in exchanging information relevant to this project.

Other Expenses

The cost for two leased lines from Santa Cruz to Stanford (35 air miles) is an annual cost and includes installation. These are needed to obtain reliable high transfer rates for interactive graphics when the graphics terminal is remote from the host. Publication costs include costs of reprints, photography, and page charges. Manuals and documentation is to buy manuals and the cost of reproducing documentation which this project generates. Maintenance is standard 8 hour provided by DEC except for the RP03 disk which is 12 hour maintenance. All postage, telephone expenses,(rental and long distance) related to research are recharged to that research by the University. Computer time on the IBM 360/40 is to cover magnetic tape utilities and miscellaneous computing related to this research. The lease of an RP03 disk drive is based on a three year lease. The disk drive will be physically located at the SUMEX facility at Stanford, but will be supported from this grant.

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.)*
W. TODD WIPKE	Associate Professor	16 December 1940
PLACE OF BIRTH *(City, State, Country)*	**PRESENT NATIONALITY** *(If non-U.S. citizen, indicate kind of visa and expiration date)*	**SEX**
St. Charles, Missouri USA	U.S.	☒ Male ☐ Female

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

INSTITUTION AND LOCATION	DEGREE	YEAR CONFERRED	SCIENTIFIC FIELD
University of Missouri	B.S.	1962	Chemistry
University of California	Ph.D.	1965	Chemistry
Harvard University	Postdoc	1967-69	Chemistry and Computer Science

HONORS Sigma Xi, 1964; Phi Beta Kappa, 1962; University of Missouri, 1962 Honors in Chemistry Distinguished Military Graduate 1962; Omicron Delta Kappa (leadership and scholarship) 1961 Pi Mu Epsilon (mathematics), 1960; Sigma Rho Sigma (Scholarship), 1959; Commendation Medal, US Army, 1967.

MAJOR RESEARCH INTEREST	ROLE IN PROPOSED PROJECT
Organic Chemistry Structure and Synthesis	Principal Investigator

RESEARCH SUPPORT *(See instructions)*

NIH GM 22990-01 Biogenetic-Like Cyclizations of Macrocyclic Polyenes	$18,109
IBM Fellowship for Krishnan Subramanian December '75 - June '76	8,000
Merck, Sharp, and Dohme Fellowship September '75 - June '76	8,000
Initial Starting grant from UC Santa Cruz (equipment only)	4,000
Ortho Pharmaceutical (Application for support for chemicals)	1,000

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.)*

1975-	Associate Professor, University of California, Santa Cruz, CA
1969-75	Assistant Professor, Princeton University, Princeton, New Jersey
1967-69	Postdoctoral Fellow in Chemistry, Harvard University, Cambridge, Massachusetts, The Synthesis of Sativene and Related Natural Products. E.J. Corey
1965-67	Automatic Data Processing and Analysis Officer, US Army Combat Developments Command, Air Defense Agency, Fort Bliss, Texas
1964-65	NIH research assistant fellowship, University of California, Berkeley
1962-65	Graduate Work on Photochemistry of Transoid Dienes, Prof. W.G. Dauben, University of California, Berkeley
1962	Research Chemist, ESSO Research Engineering Company, Baton Rouge, Louisiana (summer)
	(1972-
Co-Chairman	ACS Symposium on Computer-Assisted Design of Organic Syntheses, April 1976
Member	National Academy of Sciences Committee to Establish a National Resource for Computation in Chemistry 1974-present
Member	Chemical Abstracts Advisory Board (1969-1972)
Member	Editorial Advisory Board of Chemical Substructure Index; Editorial Board, Computers and Chemistry, Pergamon Press; Editorial Board, Journal of Chemical Information and Computer Science, American Chemical Society
Consultant	Merck, Sharp, and Dohme, Squibb, BASF
Director	NATO Advance Study Inst. on "Computer Representation and Manipulation of Chemical Information", June 1973, Holland.

Computer Experience

W.T. Wipke

Began using computers in 1962 at Berkeley, programming the IBM 7094 DCS System in FORTRAN and FAP/MAP to solve NMR analysis problems and photochemical kinetics problems. Used the SDS - 910 system for exhaustive enumeration of all tricyclic undecanes. Spent two years as an Army Systems Analyst in Air Defense Agency, Fort Bliss, Texas, IBM 7094 at White Sands, New Mexico, and IBM 360 models 50 and 65, attended IBM Systems programmer school at Los Angeles, and learned techniques of simulation and managing the development of complex programs and coordination of programming teams.

Spent two years at Harvard, working closely with Professors Thomas Cheetham and Ivan E. Sutherland in building a program to predict organic syntheses using graphics. Audited courses in graphics, and linguistics. Developed system software on PDP-1 for graphics, color display, dynamic storage management, program overlaying, list processing, 3-dimensional display, and virtual memory systems.

At Princeton he and his group completed the first stage of SECS, a program to help design stereospecific synthesis; developed GIGL, a general interactive graphics language; ALCHEM, a language for describing chemical reactions; SYNCOM, a compiler for ALCHEM; and SYMIN, a 3-D molecular model builder and a 3-D tablet.

Synthetic Experience

Synthesis of decalins and strained ring systems (with Dauben); sesquiterpenes, sativene (with Corey); and at Princeton design and execution of novel approaches to sirenin, cantharidin, palisonin. Current syntheses underway include an approach to macrocycles and a new steroid synthesis. Other work includes a study of palladium π-complexes in synthesis. Considerable synthetic experience has also been gained in six years of designing computer programs which design syntheses, and in organizing the body of chemical reactions.

Conferences (recent)

Invited Speaker: "The Rudolph Anderson Symposium on Innovations in the Methods and Tools of Synthetic Organic Chemistry", Yale University, Jan. 14-15, 1971.
Invited Speaker: "Applications of Computers in Synthesis", Stanford Symposium on "Synthesis: A Science for All Seasons", Nov. 12-14, 1973.
Invited Speaker: Symposium on Strategies in Organic Synthesis, sponsored by Societe Chimique de Belgique at the University Louvain la Neuvre, 1974.
Invited Speaker: Artifical Intelligence in Medicine Symposium, Rutgers, 1975.

PUBLICATIONS (recent relevant)

1. E.J. Corey and W.T. Wipke, "Computer-Assisted Design of Complex Molecular Syntheses", Science, 166, 178 (1969).
2. E.J. Corey, W.T. Wipke, R.D. Cramer, and W.J. Howe, "Computer-Assisted Synthetic Analysis: Facile Man-Machine Communication of Chemical Structure by Interactive Computer Graphics", J. Amer. Chem. Soc., 94, 421 (1972).
3. E.J. Corey, W.T. Wipke, R.D. Cramer, and W.J. Howe, "Techniques for Perception by a Computer of Synthetically Significant Structural Features in Complex Molecules", J. Amer. Chem. Soc., 94, 431 (1972).
4. W.T. Wipke and A. Whetstone, "Graphic Digitizing in 3-D", Computer Graphics, 5, 10 (1971).

GPO : 1969 O - 350-360

5. P. Gund, W.T. Wipke, and R. Langridge, "Computer Searching of a Molecular Structure File for Pharmacophoric Patterns," Computers in Chemical Research and Education, Elsevier, Amsterdam, vol. II (1973) pp 5/33-38.
6. W.T. Wipke and T.M. Dyott, "Simulation and Evaluation of Chemical Synthesis. Computer Representation and Manipulation of Stereochemistry," J. Amer. Chem. Soc., 96, 4825 (1974).
7. W.T. Wipke and T.M. Dyott, "Stereochemically Unique Naming Algorithm," J. Amer. Chem. Soc., 96, 4834 (1974).
8. W.T. Wipke and P. Gund, "Congestion: A Conformation-Dependent Measure of Steric Environment. Derivation and Application in Stereoselective Addition to Unsaturated Carbon," J. Amer. Chem. Soc., 96, 299 (1974).
9. W.T. Wipke,"Computer-Assisted Three-Dimensional Synthetic Analysis," in Computer Representation and Manipulation of Chemical Information, ed. W.T. Wipke, S.R. Heller, R.J. Feldmann, E. Hyde, John Wiley, (1974), pp 147-174.
10. W.T. Wipke and T.M. Dyott, "Use of Ring Assemblies in a Ring Perception Algorithm," J. Chem. Info. and Computer Sci., 15, 140 (1975).
11. T.M. Gund, P.V.R. Schleyer, P.H. Gund and W.T. Wipke, "Computer Assisted Graph Theoretical Analysis of Complex Mechanistic Problems in Polycyclic Hydrocarbons. The Mechanism of Diamantane Formation from Various Pentacyclotetradecanes," J. Amer. Chem. Soc., 97, 743 (1975).

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.)*
GRAHAM M. SMITH	Research Associate	11 November 1947
PLACE OF BIRTH *(City, State, Country)*	**PRESENT NATIONALITY** *(If non-U.S. citizen, indicate kind of visa and expiration date)*	**SEX**
Bay Shore, New York	U.S.	☒ Male ☐ Female

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

INSTITUTION AND LOCATION	DEGREE	YEAR CONFERRED	SCIENTIFIC FIELD
Adelphi Suffolk College (Dowling)	B.A.	1969	Chemistry
State University of New York at Buffalo	Ph.D.	1974	Chemistry

HONORS

Samuel B. Silbert Fellowship 1971-1972.

MAJOR RESEARCH INTEREST	ROLE IN PROPOSED PROJECT
Organic Synthesis	Development of Computer Assisted Synthesis

RESEARCH SUPPORT *(See instructions)*

Merck, Sharp and Dohme Fellowship December 1975-June '76
IBM Fellowship February 1973-December 1975

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.)*

9/73-1/74	SUNYAB Teaching Assistantship	Qual. Organic
6/71-9/73	" Research Assistantship	
6/70-6/71	" Teaching Asisstantship	Organic Chemistry
1/70-6/70	" " "	Analytical Chemistry
9/69-1/70	" " "	Freshman Chemistry
1/67-6/69	Scanner for the Brookhave National Laboratory 80 inch Bubble Chamber Group	

PUBLICATIONS

Wudl, F., Smith, G.M., Hufnagel, E.J., "Bis-1,3-dithiolium Chloride: an Unusually Stable Organic Radical Cation", Chem. Comm., 1456 (1970).

Wudl, F., Smith, G.M., "Coordination Complexes of Alkaline and Alkaline-Earth Ions II: Synthesis and Properties of Macrocyclic and Open-Chain Amino-Ethers and Their Derivatives", presented at the 164th National Meeting of the American Chemical Society in New York City, August 1972.

Green, E.A., Duax, W.L., Smith, G.M., Wudl, F. Coordination complexes of Group I and II; Potassium-O,O'-catecholdiacetate, JACS 97 6689 (1975).

Member - American Chemical Society
Member (Assoc.) Assoc. for Comp. Mach.

SECTION II – PRIVILEGED COMMUNICATION T. Wipke [redacted]

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.)
Hartmut W. Braun	Research Postdoctoral Fellow	Aug. 31, 1947
PLACE OF BIRTH (City, State, Country)	**PRESENT NATIONALITY (If non-U.S. citizen, indicate kind of visa and expiration date)**	**SEX**
29 Freudenstadt, W. Germany	German, J-1, Sept. 1976	☒ Male ☐ Female

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

INSTITUTION AND LOCATION	DEGREE	YEAR CONFERRED	SCIENTIFIC FIELD
University of Goettingen, Germany	Diplom-Chemiker	1971	Internal rotation molecules
	Dr.	1974	Theoretical and spectroscopic methods associated with internal rotation molecules

HONORS

MAJOR RESEARCH INTEREST	ROLE IN PROPOSED PROJECT
Application of computers and programming in chemistry	Research Associate

RESEARCH SUPPORT *(See instructions)*

Jan 75-Dec 75 - Ausbildungsstipendium (training and research fellowship) from Deutiche Forschungsgeimschaft, Bonn, Germany

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.)*

Princeton: 6 months of work with the synthesis program SECS

Goettingen: 3 years of experience with infrared spectroscopy, molecular mechanics calculations and quantum mechanics calculations on problems of conformation equilibria, and programming related to these fields.

1. "Die innere Rotation der Butadien-Diepoxide", Diplomarbeit, Goettingen, 1971.
2. "Theoretische und spektroskopische Untersuchungen zum Konformationsgleichgewicht des Bicyclopropyls", Dissertation, Goettingen 1974.
3. "Die Rotationsisomerie des Bicyclopropyls. II. Die gauche/trans-Isomerisierungsenthalpie und -entropie von Bicyclopropyl aus IR-Intensitaetsmessungen. Ein experimenteller Test", J. Mol. Str. 21, 391 (1975).
4. "Ueber due Bestimmumg von Isomerisierungsenthalpien und -entropien mit der IR-Intensitaetsmethode", J. Mol. Str. 21, 415 (1975).
5. "Die Rotationsisomerie des Bicyclopropyls. III. Untersuchung der inneren Rotation von Bicyclopropyl, Vincylcyclpropan, und Butadien und einiger verwandter Verbindungen mit der Kraftfeld-Methode", J. Mol. Str., sccepted for publication.

SECTION II – PRIVILEGED COMMUNICATION T. Wipke [redacted]

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.)*
Krishnan Subramanian	Postdoctoral fellow	8-8-1949

PLACE OF BIRTH *(City, State, Country)*	PRESENT NATIONALITY *(If non-U.S. citizen, indicate kind of visa and expiration date)*	SEX
Wadakancheri (Kerala) INDIA	Indian, J-1, Nov. 1976	☒ Male ☐ Female

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

INSTITUTION AND LOCATION	DEGREE	YEAR CONFERRED	SCIENTIFIC FIELD
Bombay University (INDIA)	B.Sc.	1969	Physics
Bombay University (INDIA)	M.Sc.	1971	"
Indian Institute of Science (Bangalone) INDIA	Ph.D.	1975	Chemical Information

HONORS

National Science Talent Scholar (INDIA)

MAJOR RESEARCH INTEREST	ROLE IN PROPOSED PROJECT
Computer application to chemistry	Research Associate

RESEARCH SUPPORT *(See instructions)*

IBM Fellowship - December 1975 - June 1976

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.)*

1. ALWIN - Algorithmic Wismesmer Notation System for Organic Compounds, J. Chem. Doc. 14, 130 (1974).

2. A Simplified Grammar for Algorithmic Wismesser Notation Using Morgan Name (to appear in International Classification).

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.)*
GLENN I. OUCHI	Research Associate	23 August 1949
PLACE OF BIRTH *(City, State, Country)*	**PRESENT NATIONALITY** *(If non-U.S. citizen, indicate kind of visa and expiration date)*	**SEX**
Compton, California	U.S.	☒ Male ☐ Female

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

INSTITUTION AND LOCATION	DEGREE	YEAR CONFERRED	SCIENTIFIC FIELD
University of California, Los Angeles	B.S.	1971	Chemistry
University of Minnesota	M.S.	1975	Organic Chemistry

HONORS

Kodak Summer Fellowship 1972

MAJOR RESEARCH INTEREST	ROLE IN PROPOSED PROJECT
Organic Synthesis	Development of Computer Assisted Synthesis

RESEARCH SUPPORT *(See instructions)*

Teaching Assistantship UCSC

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.)*

9/75-present	UCSC Teaching Assistantship, Organic Chemistry
1/75-9/75	Foothill College, Instructor of Chemistry, General/Organic
4/73-9/75	Research Chemist, Stanford Research Institute
9/72-3/73	University of Minnesota, Teaching Assistantship, Organic Chemistry
9/71-9/72	University of Minnesota, Teaching Assistantship, General Chemistry

Publication:

Ouchi, G.I., Spanggord, R.J., Francis, A.J., "Degradation of Lindane by E. coli", Appl. Microbiol. 29(4) 567 (1975).

RESOURCE-RELATED RESEARCH: BIOMOLECULAR SYNTHESIS

RESEARCH PLAN

A. Introduction.

1. Objective: The development of new drugs and the study of how drug/structure is related to biological activity depends upon the chemist's ability to synthesize new molecular structures as well as his ability to modify existing structures or to incorporate isotopic labels into biomolecular substrates. The long term objective of this research is to develop the logical principles of molecular construction and employ these in practical computer programs to assist investigators in designing stereospecific syntheses of complex bio-organic molecules of the type encountered in natural products and pharmaceuticals. While some progress has been made toward this objective, there is much to do. In this proposal we plan to build on the current SECS synthesis program, increasing our coverage of chemistry, increasing speed and efficiency of processing, and capitalizing on our steric and electronic perception in strategy and plan development. We plan to evaluate this program by making it available over a nation-wide network to interested health-related non-profit users. We will also explore other possible applications of the SECS program in chemistry and other applications of the program modules, for example to explore the forward-working approach to synthesis.

2. Background: Although instrumentation has dramatically improved the speed of structural analysis over the past 20 years, there has not been a significant increase in the number of reactions a chemist runs per month.[1] Execution of even rather simple synthetic schemes may require a commitment of several man-years of laboratory work. Sarett noted that speed in the laboratory will not change much and that the greatest gains will come in the area of synthetic design. He envisioned the use of computers in a backward-working analysis to generate a "synthesis tree".[1] Of course the reason computer analysis is desirable is that the computer can remember the many known chemical reactions, methodically apply them to generate a large number of unbiased synthetic routes, from which the chemist can select the best to actually execute in the laboratory. Thus he is assured of having considered all reasonable alternatives.

Research in this area began with the representation of molecular structure[2] and generation of isomers[2b,3] Later with Corey,[4] the first computer program (OCSS, later called LHASA[5]) was developed which generated synthetic schemes using the logic-oriented approach. At this time the programs were written in assembly language and considered only the connectivity of a structure, completely ignoring stereochemistry, steric hindrance, strain effects, and proximity, but still could produce some interesting carbocyclic syntheses.

In 1969 Dr. Wipke at Princeton began developing on a PDP-10/LDS-1 system in FORTRAN a new synthesis program called SECS (Simulation and Evaluation of Chemical Synthesis) to concentrate on stereospecific syntheses taking into consideration the three-dimensional structure of the target molecule.[6] Algorithms were developed for representing and manipulating stereochemistry,[7] building a three-dimensional model,[8] and analysing proximity and steric congestion from the model.[9] SECS included new areas of chemistry, hetrocyclic and protecting group chemistry.[10] (Further details on SECS are presented in section A4)

Meanwhile at Harvard the emphasis in LHASA was toward the building of sophisticated transforms (eg.,Diels Alder) which could cause generation of up to 15 step sequences of reactions to achieve a certain type of synthesis.[11] The rigidities of functional group interconversion in this early work appears

GPO : 1969 O - 350-360

to be at least partially overcome by a more recent algorithm (FGI) for up to 4 FGI's.[12] Other recent work includes further definition of ring[13] and appendage[14] strategic bonds and functional group protection.[15]

Elsewhere Gelernter at Stony Brook has written a PL-1 non-interactive batch program which works backwards from the target, but selects by itself nodes in the synthesis tree to be developed.[16] This program called SYNCHEM uses the Aldrich WLN file of available compounds as acceptalbe termination points. The representation of chemistry in SYNCHEM is less detailed than in SECS or LHASA, but more attention is given to traditional heuristic tree search methods.

Yet another approach is that of Ugi which is based on a matrix formalism.[17,18] The eventual goal of his approach is to make an break bonds in all possible "legal" ways without empirical rules. The biggest application of this type approach will probably be to discovery of new reactions rather than new syntheses. An implementation of this approach (CICLOPS) has been described, but unfortunately, no examples of results have appeared.[19] A new version called MATSYN is in progress.[20]

Hendrickson has made some significant contributions, not in the computer area, but in classifying reactions by changes in oxidation state[21] and more recently by half reactions.[22a] The latter is particularly useful for setting up a molecule to break strategic bonds. He also developed an approach to electrophilic aromatic substitution[21] and for manually analysing all the rings in a molecule.[22b] Sinanoglu at Yale has studied from a graph theoretical standpoint networks of reactions and numbers of pathways within such a network.[23] Jeff Powers at Carnegie Mellon has done some reaction path analysis in chemical engineering,[20] Howard Whitlock at Wisconsin has explored linguistics in the functional group switching problem and R.V. Stevens explored ene-reactions with a small special purpose program.[24]

Significant advances in other areas of chemical inference which have a bearing on synthesis include the DARC system of DuBois which utilizes an unusual description of structure,[25] the DENDRAL mass spec analysis[26] and the CONGEN structure generator,[27] the implication of structure from mass spectra by pattern recognition techniques,[28] the interactive graphical substructure search system of Feldmann and Heller[29] and the work in reaction documentation.[30] A NATO Advanced Study Institute[31] recently reviewed the state of the art in these areas.

A short history of SECS and recent advances by Dr. Wipke's group are presented in Section 4, Progress Report.

3. Rationale:

The central goals in synthetic design are generation of chemically valid synthetic routes to a target molecule and then selection of the "best" routes. One clearly can not attain either of these goals if he ignores the important principles relating chemical reactivity to molecular shape and configuration or if he ignores important areas of chemistry such as heterocyclic chemistry--yet these areas have been ignored. Therefore this research project is oriented to develop the ability to utilize stereochemistry in synthetic analysis, to develop strategies and new heuristics relating to stereochemistry, and to include heterocyclic and aromatic as well as carbocyclic chemistry. Stereochemistry includes both the conformation independent configurational relations (cis-trans) as well as the conformation dependent relations (steric hindrance, proximity, orientation). Since the chemist uses molecular models, the computer should know how to

GPO : 1969 O - 350-560

build and analyze 3-dimensional molecular models, recognize enantiomers, evaluate steric environment, etc. Only in this way can the computer have any reasonable chance of predicting reactivity accurately enough to permit selection of "best" routes.

An analysis of "Structures of Current Interst to the Chemotherapy Program, Sept 1971" showed 50% of the structures contained aromatic or heterocyclic ring systems, indicating the importance of this area of chemistry. Therefore, this research is also aimed at representation of the chemistry needed in the synthesis of these systems, analysis of electronic properties of such systems, and special strategies for their synthesis.

When we teach students organic chemistry, we first teach structure and nomenclature, next mechanisms and reactions, and finally strategies for synthesis. The same logical order pertains to teaching the computer. Initial priorities of this research focused on developing a general representation of stereochemistry, a capability to build 3-D models, and a representation of carbocyclic reactions including stereochemical consequences as well as the relationship between stereochemistry and reactions (eg, how to estimate steric hindrance). Although work in these areas continues, new priorities are 1) development of higher level heuristics and strategies which connect stereochemistry, symmetry, proximity, etc., with high level planning (e.g., how does one capitalize on the fact that one functional group is highly hindered?), and which interconnect carbocyclic and heterocyclic chemistry; 2) exploring efficient ways of constraining the generation of valid, but uninteresting synthetic pathways; and evaluation of the SECS program on problems by us and users, using feedback for further improvement. This research focuses on exactly those important aspects of chemistry neglected in other computer synthesis programs, but at the same time, this research will not neglect synthetic strategies based on connectivity alone. The more complete treatment of chemistry in our approach is expected to provide greater power in synthetic design, especially in complex biomolecular syntheses.

4. Comprehensive Progress Report

This proposal is a new proposal, not technically a renewal, but because considerable work has been done before under another NIH grant, to assist the reader in viewing what has been done, we present a short report of this progress beginning in 1970.

a) Original objectives: To design a program for computer-assisted design of complex organic syntheses which considered the three-dimensional nature of molecules and the important principles derived therefrom: stereochemistry, proximity, steric effects, strain energy, and stereo-electronic control. These principles are important to the logic used in the planning of a synthesis of most drugs and molecules.

Requirements

1) Program should easily communicate with any organic chemist
2) Program must "understand" the stereochemistry of a complex target molecule
3) Program must be able to build a three-dimensional model of molecule
4) Program must be able to analyze 3-D model for steric effects, etc.
5) Must have knowledge of reactions and be able to apply them when appropriate.
6) Must generate precursors with proper stereochemistry according to mechanism of the reaction applied and stereochemistry of target
7) Chemist should be able to easily add new reactions to reaction library
8) Program should be applicable to real problems of interest

4.b. Summary of Results

A general FORTRAN program (SECS) for designing stereospecific syntheses was written for the PDP-10/LDS-1 system at Princeton. This was the first program to accept standard structural diagram input with stereochemistry,[7] correctly manipulate structures according to stereospecific chemical transforms, and reconstruct valid structural diagrams for output. The internal representation facilitates symbolic recognition of enantiomeric, diasteriomeric, and isomorphic structures, and cis-trans relationships.[7] The program constructs a 3-D model[8] of the structure which can be viewed in 3-D[32] and modified in conformation using a 3-D acoustic tablet developed in this work.[33] Using this model SECS evaluates steric congestion at a reaction center which has been correlated with experimental product distribution.[9]

A language for representing reactions (ALCHEM)[34] has been developed which accommodates ab initio electron-pushing as well as empirical name reactions. SECS-II (1975) incorporated functional group protection, an initial approach at heterocyclic chemistry, and an electronic energy calculation module.[10] SECS-II was placed on the First Data Corporation timesharing system for access by any interested parties. SECS also was brought up on a UNIVAC with a GT40 graphics terminal in Strasbourg, France.

This research was moved to the University of California, Santa Cruz, and was granted an allocation of the SUMEX resource. SECS was converted to TENEX and the GT40 display system. Since then a new strategy module and a symmetry module have been under development as well as optimization of the program for remote graphics. SECS has also been used to find the rearrangement pathway to diamantane,[35] and for building many models of drugs for antileukemia pattern searching.[36]

4.c. Detailed Progress Report

First the organization of SECS-II will be described, then each of the previous requirements will be discussed as to specific progress in that area.

SECS-II Program Organization

SECS used to occupy 11 segments operating in 48 K words of memory, but now is no longer overlayed on the SUMEX system because with paging virtual memory, there is no observable advantage in overlays. SECS still uses disk for storage of chemistry files and structures, the disk being utilized as backing store for variable length dynamically allocated virtual memory (software implemented). Disk space is still a problem because of the need to store source files on-line and the need to have several versions available, but memory no longer is a problem.

Following the modules shown in Fig. 1, the investigator picks up the light pen or acoustic pen and draws in a molecule, using normal structural diagram notation including hashed and wedged lines (Figure 2). Alternatively if he is using only a teletype rather than a GT40 graphics terminal, he uses the teletype input routine to enter the connectivity, stereochemistry, atom types and coordinates (optional). At the completion of input the structure is analyzed by the perception module and a cannonical representation is generated.[7b] All graph theoretical perception is done in this module.

The 3-dimensional model builder then using energy minimization techniques creates a minimum energy conformation with the correct stereochemistry (Fig 3).[8] Then the electronic model builder calculates the pi-electron delocalization energy for any conjugated systems, especially aromatic, using Huckel MO methods which are fast. From this information certain reactivity information is infered. If the chemist has selected he then interacts with the strategy command handler to enter his strategic commands, bonds to break or not break, etc., features he will or won't accept in precursors, and cutoffs in quality of chemistry that is applied. When

GPO : 1969 O - 350-360

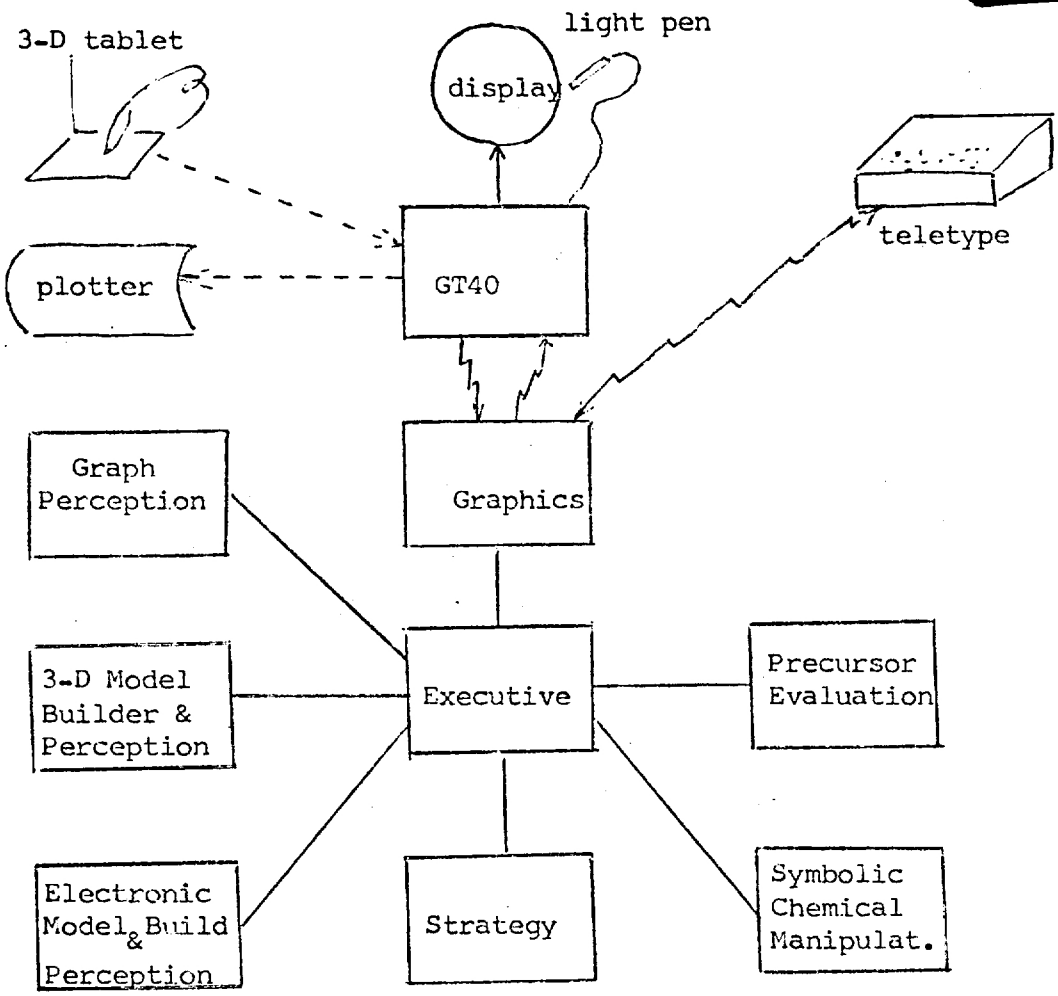


Figure 1

the investigator has specified his desires in strategy, the chemistry module attempts to find chemical transforms relevant to the structural features in the target molecule and consistent with the specified strategies. Environmental and mechanistic requirements of the transforms are examined to determine transform applicability.

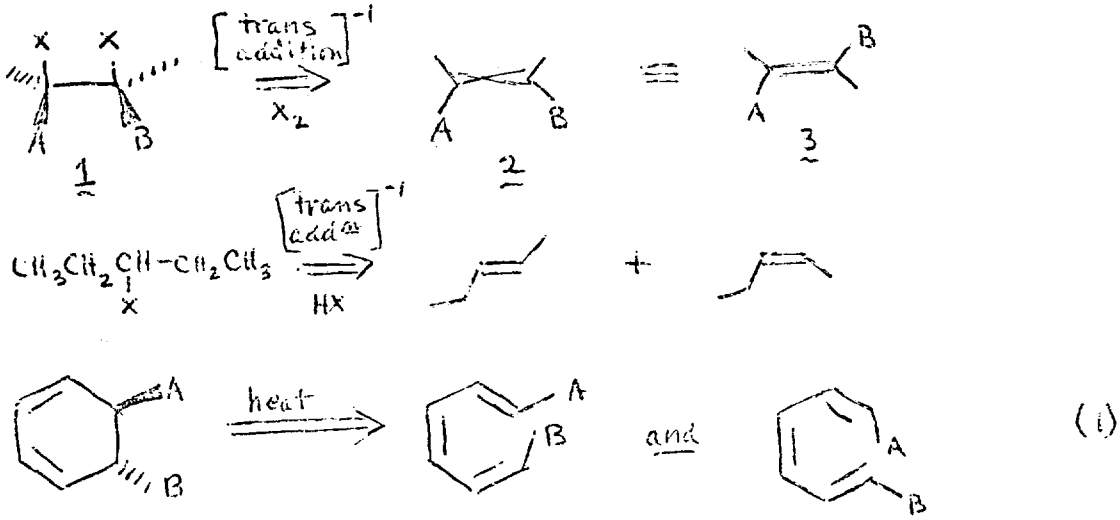
Applicable transforms generate all stereoisomeric precursors consistent with the stereochemistry of the transform and each precursor is evaluated by the evaluation module for valence violations, topologically unlikely bonding, duplications, and other undesirable features. Precursors are displayed as evaluated and are represented as a node in the "synthesis tree". Horizontal lines in the tree join ensembles of molecules needed for a transform. The chemist then evaluates the precursors, chooses one as the next to be analyzed and the process recurs. A switch causes the display of the synthetic sequence from a selected precursor up through the tree to the target at the top (Figure 4). The sequence display helps the chemist to evaluate the whole sequence since he sees the global view.

Requirement 1) easy communication of program with organic chemist. Extensive graphical communication capabilities are built into the program. Drawing of the input molecule is natural both with a tablet (best) and with the light pen(ok, but not as natural). Stereochemistry follows all natural conventions. Even the input from a teletype has been human engineered to allow the minimum of typing, eg, by describing connectivity as a path connected thru the molecule leaving only the branch points and ring closures to describe separately. The TTYINPUT module

assumes errors will occur and disallows any illegal descriptions, and provides editing capabilities for correcting errors. Output is also graphical on a teletype but of course the plotting is less attractive. However, all chemically important information is there including stereochemistry. Changes are still being made to the program to optimize the use of remote graphics terminals over medium speed lines. Perhaps the best evidence of progress in this area is that chemists at Squibb were able to use SECS on the FDC timesharing system with a GT40 terminal without any instructions other than the one page description on the system.

Requirement 2) Program must "understand" stereochemistry. We have developed an algorithm by which the computer can interpret a standard stereochemical structural diagram (Fig 2) or a 3-D model[7] and generate an identifier which is unique for each stereoisomer and allows easy recognition of enantiomeric structures which in the synthesis of racemic materials are treated as being isomorphic.[7] The algorithm ignores centers which because of symmetry are not true stereo centers.

Relative stereochemical relationships (cis-trans) are derived by another algorithm operating on the individual stereo center descriptions. The chemical manipulation module, in addition to making and breaking bonds, also generates the correct stereochemical descriptor for the precursor based on the mechanism of the operating transform. Thus 1 implies by a trans addition mechanism that the precursor could have been 2 which rewritten is 3, whereas 4 implies by the same mechanism both 5 and 6. Similarly SECS produces all valid precursors in electrocyclic transforms. (1) Note that this is not simply permutation of all possible



(1)

double bond isomers. SECS also has an algorithm to generate proper hashed and wedged bonds to correctly represent the actual stereochemistry of the precursors, even if the chemist moves atoms or rotates the molecule--the hashing/wedging is changed to maintain an accurate representation.

Requirement 3) Must be able to build a three-dimensional model. SECS contains a model-building module[8] which creates a reasonable model given only the standard two-dimensional structural diagram. It accomplishes this by special minimization techniques in an implementation of the Westheimer method[37] using four levels of parameters. Starting from any geometry, two-dimensional (Z=0), or 3-D, or even random coordinates, the program reshapes the structure into a reasonable conformation, displaying the model and strain energy as it proceeds. Initially, the program emphasizes non-bonding effects, later bond lengths, then bond angles, and finally fine resolution non-bonding factors.[8] Models for structures having up to 30 non-hydrogen atoms, double and triple bonds, and hetero-atoms may be built. During the building process SECS monitors the stereochemistry of the model and modifies the model to make it correspond to the stereochemistry initially specified.

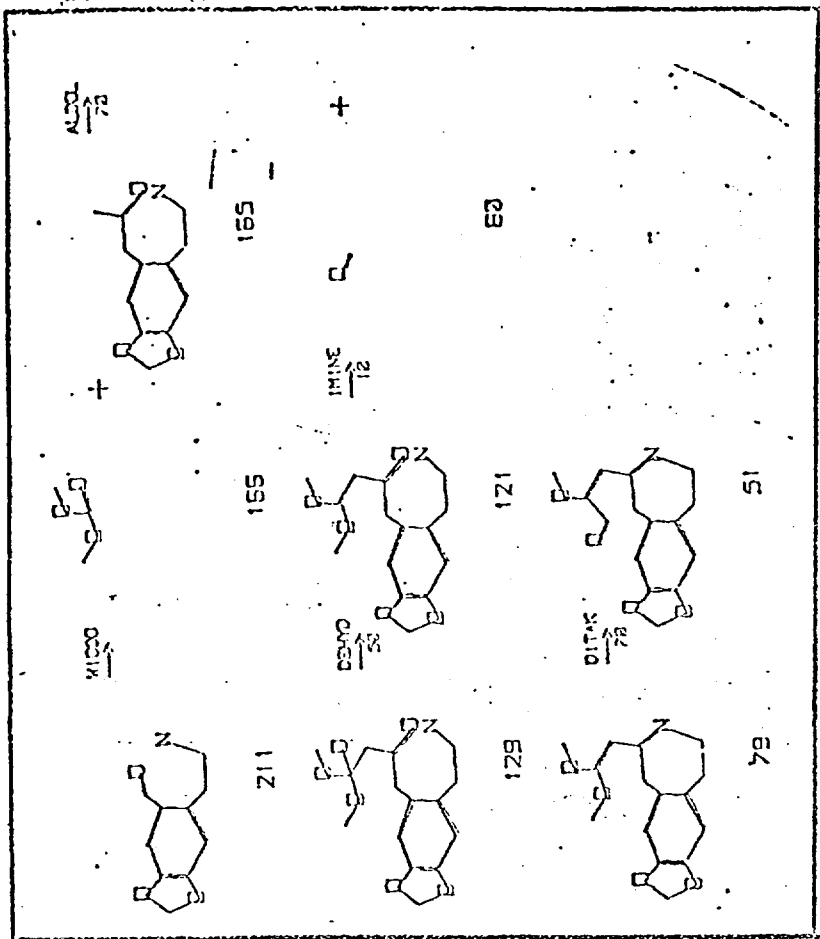


Fig. 4

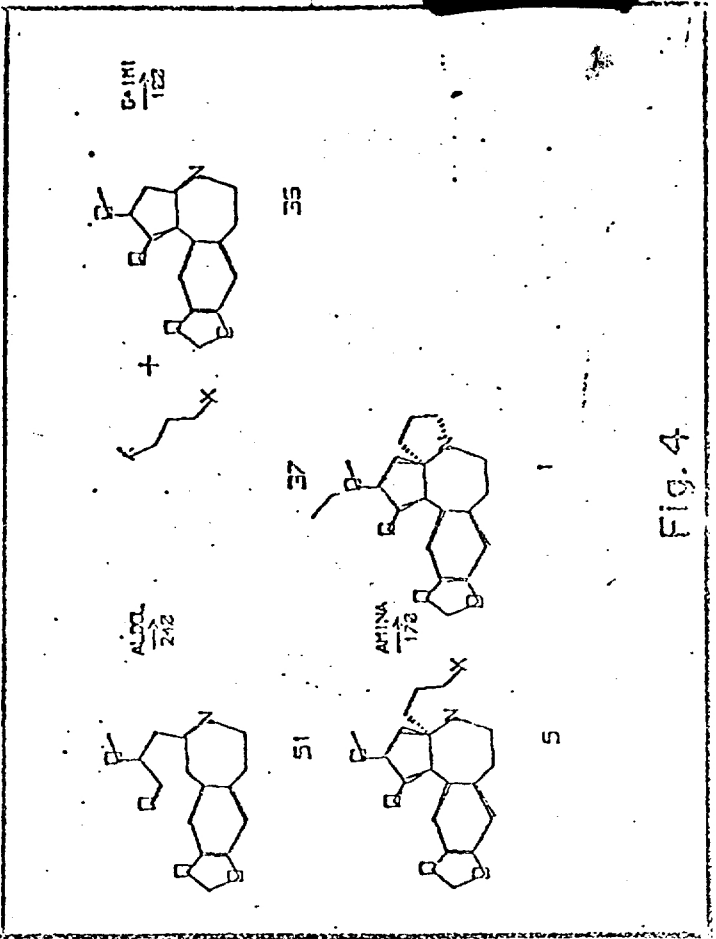


Fig. 5

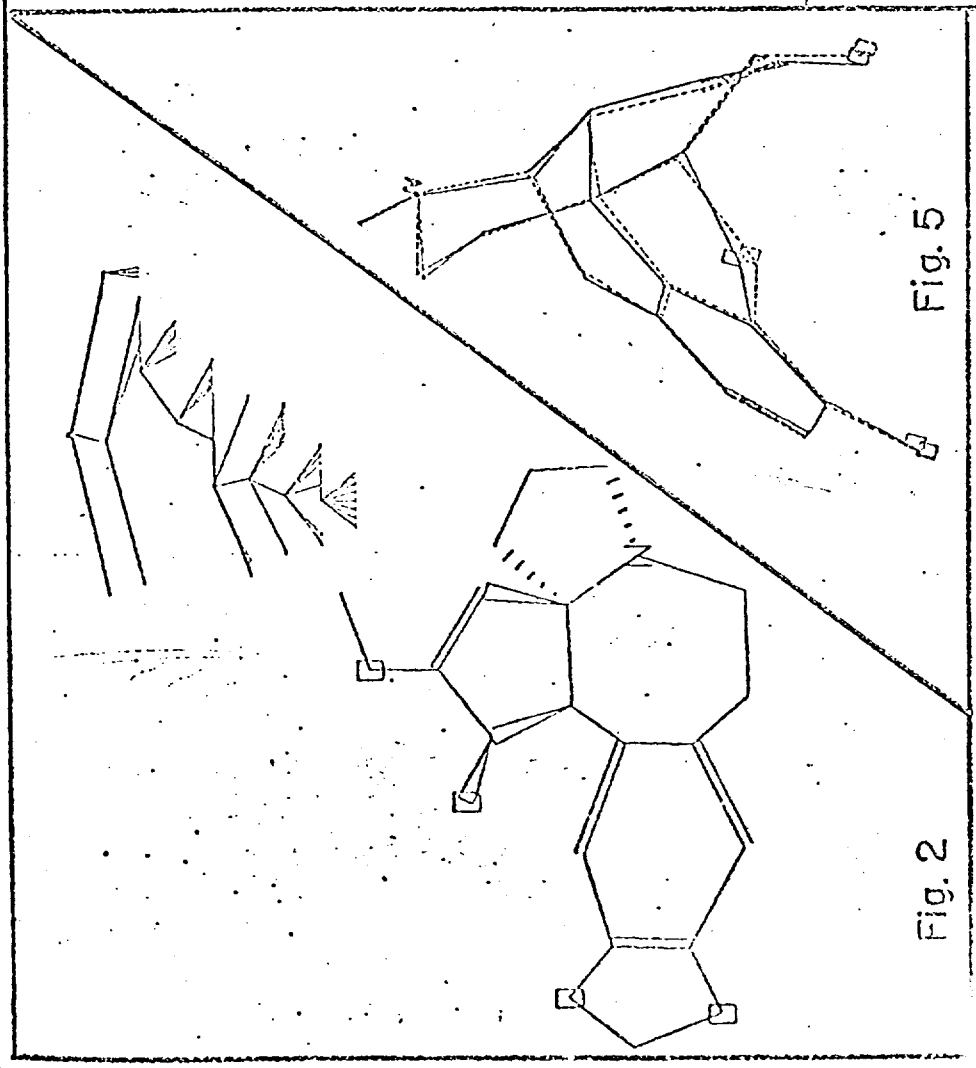


Fig. 2

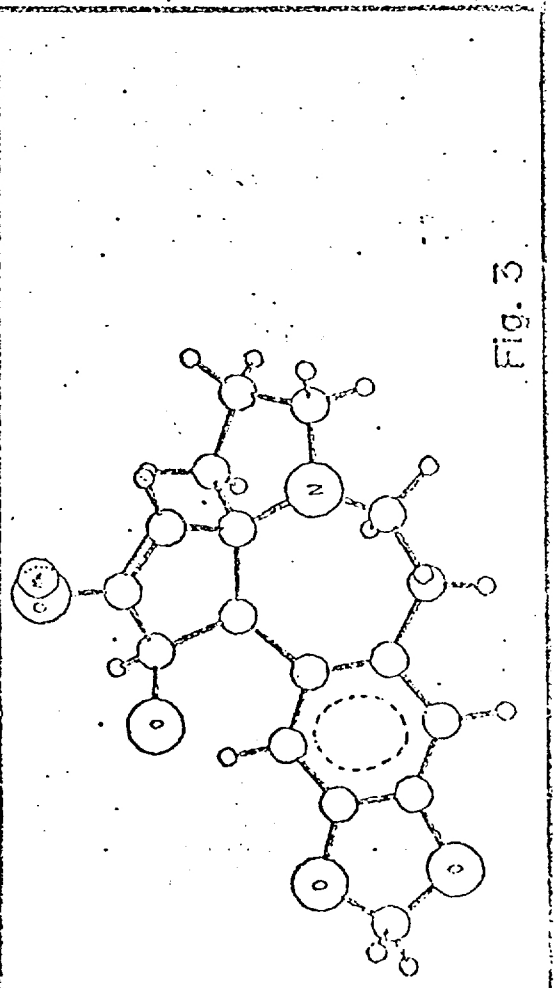
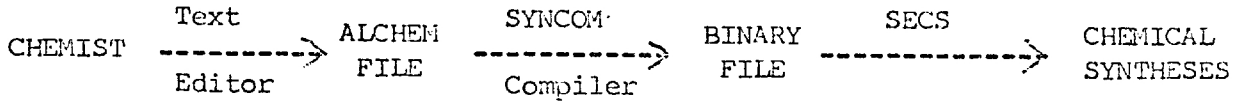


Fig. 3

Figure 5 shows a comparison of the model built for morphine (dotted line) and the x-ray structure of morphine methiodide salt (solid line). This model builder has been provided to Feldmann at NIH, is incorporated into the PROPHET pharmacology information system, and into the CONGEN-DENDRAL system at Stanford. Peter Jurs, (Penn State) and Bruce Kowalski (U. of Washington) also have been provided a copy of the model builder for use in pattern recognition studies. Squibb and Pfizer have used the model builder on the FDC system.

Requirement 4) Must be able to analyze 3-D model for steric effects. Certain ALCHEM statements cause the synthesis program to access it's internal 3-D model much as a chemist would do, making distance and angle measurements for proximity and syn or anti relationships. Evaluation of steric hindrance is considerably more difficult since one must first define steric hindrance. Using collision theory, we have developed a definition of ground state steric congestion at a reaction center assuming the reacting partner is an infinitesimal particle.[9] This function works well for rigid ketones where there is rather large congestion. However for rather uncongested ketones, we had to include an electronic eclipsing effect based on the dihedral angle the incoming group makes with substituents attached to the alpha carbons of the ketone. When these two functions are combined they provide a quantitative treatment of steric hindrance not only for reduction of ketones, but also epoxidation of olefins. Since our functions give an absolute value for attack to each side of a planar group, we can also compare the least hindered side of the two identical groups in a target to determine the feasibility of carrying out a selective reaction on one of the groups in the presence of the other. Currently the only chemistry using this steric information is the reduction of ketones and Grignard reactions on ketones, but the correlation of congestion with product ratios in these cases is very good.

Requirement 5) Must have knowledge of reactions and be able to apply them when appropriate. Chemical transforms (a transform can be a simple ab-initio electron pushing step or an empirical name reaction written in the antithetic direction) are written in an English-like chemical language, ALCHEM. See appendix for the BNF grammar of ALCHEM. Basically ALCHEM has facilities for describing relationships between functional groups, structural features, atoms, and bonds, and even arbitrary substructural fragments. It also contains general arithmetic capability



and means for symbolically manipulating structures to generate precursors. Since ALCHEM files are ASCII text and the compiler (SYNCOM) and interpreter (SECS) are in FORTRAN, we have a machine-independent language for describing reactions and the factors affecting them. Any reaction may be described in any amount of detail with the only limit to the number of transforms being the amount of disk space available.

Requirement 6 has already been discussed under item 2.

Requirement 7) Chemist should be able to easily add new reactions to reaction library. All that is needed to add a new reaction is to type it into the text editor in ALCHEM, adding it to one of the existing files or creating a new one; then run SYNCOM to take the ALC file and create the binary CHM file; then just run SECS and it will automatically use the new updated CHM file. Total time 5 min. Dr. Guenter Grethe entered many complex reactions without knowing any programming languages, only ALCHEM and how to operate the text editor, so it is possible for a chemist to easily enter reactions. It took him about a day of study of ALCHEM.

Requirement 8) Program should be applicable to real problems. SECS was designed to handle molecules of up to 72 non-hydrogen atoms, and to be able to generate any number of structures in the synthesis tree. It is able to do this because the structures are not in memory, but are in our own software implemented virtual memory which can be saved and restored in another session to allow interruptions. A version of SECS on FDC timesharing system in Boston has received considerable testing by Squibb, Pfizer, and Merck, Sharp and Dohme pharmaceutical companies in the past year. The feedback from these users is that they were giving it real problems and were finding interesting output which they had not thought of. This is not to say that everything produced was good or that every good route was produced, but it does say some progress has been made on real problems.

Figure 2 shows the antileukemic cephalotaxine[38] as a target with the edited tree. Figure 3 shows the model created and figure 4 shows one of the routes formated by the synthetic sequence layout. Below each structure is its sequence number. Above each arrow is the code name for the transform implied and below that is the final priority ranking of that transform. This route is different from an earlier attempted synthesis[39] and recent successful syntheses.[40]

Preliminary Results

Aromatic and heterocyclic chemistry. Our initial work in this area uses the powerful pattern transform capability of ALCHEM. We have about 100 heterocyclic transforms which represents many times more reactions and fairly well covers the simple ring systems to a first approximation. Electrophilic substitution directing effects are included, also steric effects from groups already on the ring, but only for rings not containing heteroatoms. Recently we generalized this using the new electronic energy model with fairly good success. This research showed the difficulties with tautomerism (keto-enol) and with strategic control of when to perform the synthesis of the aromatic or heterocyclic ring system.

Functional Group Protection. Dr. Willi Sieber created a module for checking the condition statements in the transform and automatically invoking protection, selecting the proper protecting group which would be stable to the conditions yet not react with other groups in the molecule. This protecting group is a text descripter which is attached to the functional group for that step. Corey just published a similar approach.[15]

Strategy Research. A general goal list structure with an interactive creation package now allows us to manually specify trial strategies involving not only the breaking and making of bonds, but also selection of transforms by character, e.g., rearangement, ring closure, modify stereochemistry. This has increased selectivity and suggested ideas for more advanced specifications to control sequences of reactions. The strategy of striving for "simplicity" has been explored by Peter Friedland by letting the executive choose the next structure to be processed on the basis of a simplicity function. This function depends on the size of the molecule, number and size of rings, appendages, functional groups, and ring junctures. Stereo and group sensitivity were also incorporated. The user was allowed to set the depth of the search and the program processed the "simplest" structure in each set of precursors produced. For some syntheses this works well, but we find for may others that the computer does not find very interesting syntheses by itself because the interesting syntheses must go through a more complex intermediate structure.

Symmetry. Work is underway to detect the symmetry of a target molecule and use this to prevent redundant reactions due either to molecular symmetry or symmetry of the reaction. This is related to the use of symmetry that the CONGEN program,[27b] but is different because of the fact we are mapping reactions rather than just a set of labels, and also because we are including stereochemistry in

GPO : 1969 O - 350-360

our molecules and our symmetries in most cases correspond to point group symmetry elements. For example, for dodecahedrane we find 124 symmetry elements, the same as the point group for the molecule. We expect this symmetry information to be very useful in not only reducing the number of precursors created to a small number of unique precursors, but also useful in building strategic goals. We expect that we may also be able to use this symmetry to prevent the generation of enantiomers which are not normally desired unless one is dealing with resolved materials. In some cases, symmetry information will reduce the execution time of SECS by a factor of five or more.

Conclusion

Many problems still remain. We are just beginning to learn how to use the powerful perceptual information we have collected, but we feel we have established a firm foundation on which to build further research in this area. See letters in Appendix 1.

4.d. Publications

W.T. Wipke and A. Whetstone, "Graphic Digitizing in 3-D," Computer Graphics, 5 (4), 10 (1971).

W.T. Wipke, "A New Approach to Computer Assisted Design of Organic Syntheses," Proceedings of Northern Illinois University Conference on Computers in Chemical Education and Research, DeKalb, July 19-25, 1971, p. 10-60.

E.M. Engler, L. Chang and P.v.R. Schleyer, "The Flexibility and Conformations of Polycycloalkanes with Two-Carbon Bridges," Tetrahedron Letters, 2525 (1972).

T.M. Gorrie, E.M. Engler, R.C. Bingham, and P.v.R. Schleyer, "An Abnormally Weak Bond in a Diol with a 0° Dihedral Angle. Breakdown of the OH ... OH Spectral Shift-Dihedral Angle Relationship," Tetrahedron Letters, 3039 (1972).

W.T. Wipke and P. Gund, "Congestion: A Conformation-Dependent Measure of Steric Environment. Derivation and Application in Stereoselective Addition to Unsaturated Carbon," J. Amer. Chem. Soc., 96, 299 (1974).

W.T. Wipke and T.M. Dyott, "Simulation and Evaluation of Chemical Synthesis. Computer Representation and Manipulation of Stereochemistry," J. Amer. Chem. Soc., 96, 4825 (1974).

W.T. Wipke and T.M. Dyott, "Stereochemically Unique Naming Algorithm," J. Amer. Chem. Soc., 96, 4834 (1974).

W.T. Wipke, "Computer-Assisted Three-Dimensional Synthetic Analysis," in Computer Representation and Manipulation of Chemical Information, ed. W.T. Wipke, S.R. Heller, R.J. Feldmann and E. Hyde, J. Wiley, (1974) pp 147-174.

P. Gund, W.T. Wipke, and R. Langridge, "Computer Searching of a Molecular Structure File for Pharmacophoric Patterns," Computers in Chemical Research and Education, Elsevier, Amsterdam, vol. II (1973) pp 5/33-38.

W.T. Wipke and T.M. Dyott, "Use of Ring Assemblies in a Ring Perception Algorithm," J. Chem. Info. and Computer Sci., 15, 140 (1975).

T.M. Gund, P.v.R. Schleyer, P.H. Gund and W.T. Wipke, "Computer Assisted Graph Theoretical Analysis of Complex Mechanistic Problems in Polycyclic Hydrocarbons. The Mechanism of Diamantane Formation from Various Pentacyclotetradecanes," J. Amer. Chem. Soc., 97, 743 (1975).

GPO : 1969 O - 350-360

4.e. Staffing

W. T. Wipke, Ph.D.	Associate Professor	4/70-
P. Gund, Ph.D.	Postdoctoral	9/70-10/73
C. Still, Ph.D.	Postdoctoral	3/72-6/73
T. Brownscombe, Ph.D.	Postdoctoral	9/72-10/73
G. Smith, Ph.D.	Postdoctoral	2/74-
S. Krishnan, Ph.D.	Postdoctoral	12/75-
F. Choplin, Ph.D.	Visiting Fellow	9/73-10/74
H. Bruns, Ph.D.	Visiting Fellow	1/72-4/72
G. Grethe, Ph.D.	Visiting Fellow	9/72-8/73
M. Spann	Visiting Fellow	6/73-9/73
W. Sieber, Ph.D.	Visiting Fellow	1/74-12/74
H. Braun, Ph.D.	Visiting Fellow	1/75-
T.M. Dyott, B.S.	Graduate Student	4/70-8/73
D. Stevens, B.S.	Graduate Student	7/72-8/72
S. Stevens, B.S.	Graduate Student	3/72-8/72
G. Goeke, B.S.	Graduate Student	3/72-8/72
J. Mitlitzky, B.S.	Graduate Student	9/72-1/73
G. Ouchi, M.S.	Graduate Student	1/76-
C. Marikakis, B.S.	Graduate Student	7/72-8/72
T. Su, B.S.	Graduate Student	7/73-1/74
P. Friedland	Undergraduate	9/70-9/74
J. Verbalis	Undergraduate	4/70-6/71
J. Jackson	Undergraduate	9/71-6/73
A. Zelicoff	Undergraduate	6/73-6/75
T. Newman	Undergraduate	9/75-
D. Shapiro	Undergraduate	9/75-
T. Davis	Undergraduate	9/75-

B. SPECIFIC AIMS

Our objective is to increase the speed, efficiency and reasoning power of the Simulation and Evaluation of Chemical Synthesis program, capitalizing on the steric and stereochemical information from our previous work. Specific aims for this project period are listed below according to the module in which they fall.

1. Symmetry
   a) Develop an efficient molecular symmetry recognizer using stereochemistry
   b) Incorporate symmetry into transform applicator to eliminate generation of redundant precursors and enantiomers
   c) Investigate algorithms for detection of potential symmetry
2. Model Builder
   a) Generalize to more diverse types of bonding
   b) Increase speed and reduce calculation using heuristics for appendages, symmetry, etc.
3. Strategy and Planning
   a) Continue to explore and evaluate principle of separation of strategies from chemical transforms
   b) Develop strategy modules for steric, proximity, symmetry, and electronic factors
   c) Evaluate special strategies for heterocyclic and aromatic chemistry
   d) Build a graphical interface to the strategy executive module

4. Chemical transforms

a) Fill in and extend transform library in heterocyclic and aromatic area
b) Develop interactive programs for update and maintenance of library
c) Explore additional methods for acquiring information from users
d) Expand capability to predict chemical reactivity from 3-D models

5. Evaluation

a) Develop forward-working simulator for plan evaluation
b) Evaluate feedback from users

C. METHODS OF PROCEDURE

Our current synthesis program (SECS-II) will serve as the foundation for future investigations in synthesis planning. It should be clear that the list of specific aims given above is not exhaustive, but illustrative, and is our view at this point in time. Each of these aims will now be discussed in more detail.

Symmetry. The most obvious need for symmetry is simply to prevent the generation of redundant precursors. Currently if SECS infers that cyclohexane could come from cyclohexene, since there are six single bonds that could be changed to a double bond, six cyclohexene precursors would be generated, then five would be deleted! Clearly it is more efficient to only produce one precursor in the first place, knowing that all bonds are equivalent. Work is underway to develop an efficient molecular symmetry recognizer, using graph theory and our representation of stereochemistry. The validity of this algorithm will have to be proven and tested. This symmetry information will then be incorporated into the various chemical transform applicator modules so that transforms are mapped onto the structure only in unique ways.

We also plan to use symmetry to constrain the generation of enantiomers since in some cases this needlessly doubles the number of intermediates generated. In the synthesis of racemic compounds, we treat enantiomers as being identical. Symmetry will be used in strategy to find ways to cleave the target into identical fragments, and also in a higher sense to prevent redundant strategies. Later we plan to investigate algoritms for detection of potential symmetry. While a general solution to this problem is not yet apparent, for certain classes of potential symmetry, the problem may be more tractible, e.g., when one must break one or more bonds to obtain a symmetrical structure. Symmetrical here means having more operators in its symmetry group than just the identity operator.

Model builder. Another goal is to improve the model builder (SYMIN) which calculates the minimum energy and optimum geometry of a chemical structure on the basis of classical mechanics. SYMIN should be generalized to handle more types of bonding, including hydrogen bonding, and electrostatic effects which are currently ignored. There is also a need to make it as efficient as possible. The symmetry information discussed above may be a useful heuristic in this regard. Other heuristics can also be used, e.g., recognizing appendages as units, staggering them and then moving them as a unit. Large appendages minimize slowly in the atom-by-atom algorithm. Another heuristic approach is to apply weighting factors to each atom so that in early stages peripheral atoms are free to move larger distances in each iteration. The numbering from our Stereochemically Extended Morgan Algorithm (SEMA) are an interesting set because they map a spanning tree onto the molecule with the origin of the tree usually corresponding to the center of gravity of the molecule.

It would also be useful if the chemist could impose general constraints on SYMIN to force the model to meet certain requirements. Ivan Sutherland's Sketchpad system had such general constraints but in a different problem area. These improvements to SYMIN would be useful to the DENDRAL CONGEN program and the PROPHET

pharmacology system, both of which use SYMIN, as well as to SECS.

Strategy and Planning. The simplest strategy is just to apply all transforms that fit--analogous to the chess legal move generator. This is a useful stragegy for exhaustively exploring a limited region of a problem, but normally it is desireable to constrain building the synthesis tree according to various other strategies. Unlike game playing, where a look-ahead can save one from making a bad move now, in SECS doing the look-ahead is the same as 'playing the move,' in that we have to do the expensive part--the chemistry--and there is no way of recovering the effort if it proves to be a fruitless direction. Deleting bad precursors is merely a way of hiding our mistakes.

A major area of interest is the development of heuristics to guide the search for simplification of the synthetic problem. Topological heuristics have already been implemented. We are now in a unique position to develop heuristics based on symmetry, stereochemistry, spatial orientation, strain energy and electronic factors. There are two important problems, 1) actually devising the heuristics and 2) actually implementing them in a synthetic program. Let us first examine the latter problem and then turn to what the heuristic might be like.

Rather than building strategies into the transforms as complex transforms consisting of preplanned sequences of subroutine calls (e.g., Corey's Diels Alder transform[11]), our approach is to try to cleanly separate strategy from transforms. We feel this is important to allow transforms to be added or updated without modifying the strategies and vice-versa. As the base of reactions increases, the importance of this separation becomes more obvious. A large library would be difficult to maintain otherwise. Another advantage of this approach is that goals can be ordered on merit rather than on the order of the statements in the transforms. Strategy modules will vote on strategic operations, e.g., breaking various bonds, or making certain bonds. The more votes a given goal has, the more powerful it becomes to demand subgoal creation. After all strategic modules have voted, the chemistry modules are passed once to satisfy these goals at a given level, rather than making multiple passes as is required when strategy modules directly call transforms.[14] This approach promises the least bias, the most creative results, but still controls the amount of output to within reasonable limits.

The planner will create short and long range plans, the latter extending over many levels of the synthesis tree, e.g.,"it is desired that a portion of the molecul remain untouched for most of the analysis." The chemist will be able to interact with the planner graphically or via teletype to modify or create additional plans.

Strategies based on Steric Effects. Steric congestion about a center can now be approximated by a function for ketones and olefins, and we hope to be able to extend it to SP^3 centers for oxidation, substitution, and elimination reactions. Congestion should be helpful in recognizing natural functional group selectivities, or adverse reactivities, and may help determine the priority for constructing stereocenters. Heuristics can be developed such as "work with the most accessible groups first, and later use the sterically congested ones." Sophisticated goals for reconnections and special blocking groups can be keyed by a need for greater congestio on one side of a reaction center.

Proximity is currently available from the 3-D model. Other proximity effects, e.g., anchimeric assistance, depend upon the orientation of a bond as well as the distance between groups. Proximity will also lead to recognition of differential reactivity and would key operations like congestion strategies.

Strategies based on Electronic perception. Currently the prime application is viewed as directing effects and relative reactivities on aromatic substrates. We currently have a molecular orbital module in SECS to provide needed perception. Related to this is the use of strain energy to trigger the need for mild irreversibl

synthetic methods. This also may focus attention on exactly what the trouble spots in a synthesis might be so that aspect may be dominant in planning.

Related to aromatic and heterocyclic systems is the problem that many different ring systems occur, all having different chemical properties. The use of HMO calculations eliminates the need for tables of data for each ring system. Of course there are limitations to the method, especially as the number of heteroatoms increases, but it provides a valuable heuristic in planning. More accurate evaluations can be applied once the synthetic routes of choice have been selected for refinement. More general strategies are needed however to determine when in a synthesis it is appropriate to work on the aromatic ring, and when it is appropriate to work on the remainder of the molecule. We are currently collecting data on this very question. Without this strategy, the behavior of the program is to jump around the molecule working here, then there, then here again, which is not the approach normally used in synthesis.

A graphical interface between the chemist and strategy module is planned to make the communication faster and more natural. There will still be complete capability also available from just a teletype for those users without CRT terminals.

Chemical Transforms. The ALCHEM libraries should be completed and extended in aromatic and heterocyclic chemistry. As the transform libraries continue to grow, finding out whether a certain reaction is present in the file becomes an increasing problem. Thus there is a need to develop programs for the maintenance and updating of the libraries, both to ease the problem mentioned and to assure consistency in the files. At the same time an interactive graphical interface to the chemist could be incorporated so he could inquire or enter a transform graphically. We would also like to generalize the ALCHEM representations so they could be used not only for the backward-working SECS, but also for retrieval and for a forward-working synthesis simulator.

The question arises of what SECS should do when a needed transformation isn't in the library. One approach is to let the chemist manually perform the transform at that stage in the analysis, then let SECS continue. SECS could remember that transform then, providing yet another way to extract information from users. Another approach is to let SECS revert to mechanistic generation at this gap to try to "invent" the desired transform.

In order to give transforms additional specificity based on our 3-D model, we plan to add to ALCHEM geometric descriptors defining lines, planes, angles, and other relationships. The arithmetic capability of ALCHEM would then allow manipulation and evaluation of these parameters. These capabilities will also be useful in representing biological reactions such as biosyntheses or metabolism. Other extensions to ALCHEM will include specification of short range "block avoidance", e.g., recognizing situations which prevent the use of a transform, automatically creating a subgoal to circumvent or correct the situation. This would be stated as IF . . . THEN CORRECT. If the situation could be corrected, the original transform would again be attempted, else it would fail. And finally, we will continue our research to correlate chemical structure and chemical reactivity as we have done with steric congestion.

Synthesis Plan Evaluation and Optimization. SECS is a planning program which works primarily backward and necessarily in generalities. It produces a synthetic plan (tree) with transform names, structures, and indications of protection and conditions. The next step is to develop a second program to perform the transforms in the forward direction, and evaluate factors more accurately than could SECS. It could also optimize reaction sequences, determine possible simultaneous steps and determine more accurately the requirements of protecting groups, when they must be introduced and removed. (Currently all that is specified is that the protecting group must be present in a particular step.)

GPO : 1969 O - 350-360

For the selected sequences, reactions and exact reagents would be specified, either by the chemist or by the program using a reaction data base. It could predict possible side products from the specified reaction, could retrieve appropriate literature references for experimental procedure, and even could label those structures in the tree that would be new compounds, not present in Chemical Abstracts Registry system.

Although the subprograms of SECS-II would be useful in the forward working program, the organization will be quite different. Ignoring the product shown in the synthesis tree obtained from SECS, the forward worker will attempt to predict the expected products, given the reactants and conditions. This involves finding reactions which can occur between the groups present under those conditions, estimating the relative extent of reaction and ratio of stereoisomers. It must of course consider both intra and inter-molecular reactions. The quality of the predictions will be dependent on the quality of the simulations and evaluation of the many factors involved. The description of the reaction mechanism would be in an ALCHEM-like language, but would view transforms from the side of reactants and conditions.

Evaluation by users. Feedback from users of SECS during this period will assist in improving the program, the approach, and in planning the research. Input of chemical transforms as well as intersting problems is expected from users. The number of users permitted and selection of them as well as the procedures involved will be worked out with the SUMEX-AIM community and administrators. It is also planned to explore other applications of SECS to demonstrate the generality of this program as a chemical problem solver. A specific problem of current interest is application in metabolism of drugs.

D. SIGNIFICANCE

This proposed research in computer-assisted biomolecular synthesis should enable investigators to plan more efficient syntheses faster, and with assurance that all the important routes have been considered in a methodical and unbiased way, using all reactions available. Consequently, the synthesis of molecules of importance to the national health care program should be achieved ultimately faster and at lower cost. This has direct bearing on the synthesis of drugs, labeled compounds for testing, and structure proof by synthesis. The many NIH grantees working on synthetic projects could gain access to this tool for assisting them in their own research program.

Further, the applications of advanced computer science to chemistry in this research has important implications for the solution of other important health-related problems involving molecular structure and man-machine communication. For example, molecular symmetry involving stereochemistry as we will find it would be required for prediction of C-13 or proton magnetic resonance spectra. Such symmetry information will be useful in generating all possible stereoisomers in the CONGEN structure generator program. The model builder is useful for generating geometries of molecules to predict spectra, or for biological activity comparison. That program is incorporated into the PROPHET pharmacology system now and would be improved by this research. And the SECS program itself may also prove useful in analysis of metabolism pathways which could be indicative of potentially dangerous metabolites of drugs or other foreign substances. Our concerns with managing a large growing data base are similar to those involved with medical diagnosis programs and drug prescription programs like MYCIN. Our analysis and evaluation of chemical reactivity from three-dimensional structure has potential applications in analysis of structures for biological activity according to shape, and has also importance in chemistry in understanding how reactions really occur and how reactic are sensitive to their environment. Lastly, our work in interactive computer

graphics has and will continue to establish techniques for efficient man-machine interaction which are useful to many other computer applications in medicine.

The significance of this research to the SUMEX-AIM community is that it provides an extended capability for dealing with biomolecular structures, for representing chemical transformations, and for evaluation of properties of biomolecules based on their graph theoretical and three-dimensional structure. This project will serve as an experiment for the feasibility of remote interactive graphics which is of interest to many SUMEX-AIM research projects. Our experience and research results will be available to all of the community, as well as our programs and algorithms. In this past six months of being a part of the SUMEX-AIM community, I believe this research has benefited from the dynamic interaction with other scientists interested in advanced computer science applied to medical problems. Our joint group meetings with the DENDRAL project at Stanford has been stimulating and hopefully will be able to continue. Already our model builder and graphics techniques have been adopted into SUMEX-AIM research programs of other investigators. I am sure there will be a continuing transfer and sharing of science and technology in the future.

SUMEX-AIM is very important to this research, for it is the only source of computing available which meets the needs for large interactive programs like SECS. Thus, this project as it is presently conceived is dependent on the availability of a large sophisticated timesharing system. Since there is none at Santa Cruz, access to the SUMEX system is important to the success of this research.

E. FACILITIES AVAILABLE

1 DEC GT40 graphics terminal
1 CDI model 1030 thermal printing terminal
1 9 trk high speed magnetic tape drive
IBM 360/40 computer (computer center)
500 square feet of space for this project, equipment and desks

F. COLLABORATIVE ARRANGEMENTS

As mentioned above, this project requires access to the SUMEX Resource at Stanford. The author is currently actively participating in the SUMEX Resource and has been for the past 6 months since he moved from Princeton to Santa Cruz. Currently SUMEX is providing computer time, disk space (4 K pages), one leased line, modems for that line, and TYMNET access. The author has discussed this proposal with NIH and with Dr. Joshua Lederberg, principal investigator of the SUMEX resource. The arrangements we have made are that this project would attempt to cover all terminal and communication costs, would provide a disk drive to increase the file space on SUMEX which is desperately needed, and SUMEX would provide an allocation of computer time and disk space for this research. A letter from Dr. Lederberg describing these negotiations is enclosed as Appendix 1.

G. PRINCIPAL INVESTIGATOR ASSURANCE

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

2 5 76 76 — Ted Wipke

Date — Principal Investigator

DO NOT TYPE THIS SPACE-BINDING MARGIN

GPO : 1969 O - 350-360

References.

1. Lewis H. Sarett, "Synthetic Organic Chemistry: New Techniques and Targets," presented before the Synthetic Organic Chemical Manufacturers Association, June 9, 1964.
2. a) L. Spialter, J. Amer. Chem. Soc., 85, 2012 (1963); J. Chem. Doc., 4, 261, 269 (1964);
   b) J. Lederberg, "Dendral 64 - A System for Computer Construction, Enumeration, and Notation of Organic Molecules as Tree Structures and Cyclic Graphs Part II Topology of Cyclic Graphs," December 1965 CR-68898, Star No. N66-14074.
3. W.T. Wipke, All tricyclic undecames were enumerated by computer, unpublished work, 1964.
4. E.J. Corey and W.T. Wipke, Science, 166, 178 (1969); W.T. Wipke, Computer Decisions, 2, 39 (1970); Sci. Amer., 222. 72 (1970).
5. E.J. Corey, W.T. Wipke, R.D. Cramer, and W.J. Howe, J. Amer. Chem. Soc., 94, 421 (1972) and adjacent papers; E.J. Corey, Quart. Rev., 25, 455 (1971).
6. W.T. Wipke, "A New Approach to Computer Assisted Design of Organic Syntheses," Proc. Conf. Computers in Chemical Education and Research, Northern Illinois, July 19, 1971, p 10-60.
7. a) W.T. Wipke and T.M. Dyott, J. Amer. Chem. Soc., 96, 4825 (1974);
   b) W.T. Wipke and T.M. Dyott, J. Amer. Chem. Soc., 96, 4834 (1974).
8. W.T. Wipke, P. G nd, T.M. Dyott, and J. Verbalis, "A Generalized 3-Dimensional Molecular Model-Building Program," presented at the 7th Middle Atlantic Regional Meeting of the Amer. Chem. Soc., Phila., Feb. 14-17, 1972.
9. W.T. Wipke and P. Gund, J. Amer. Chem. Soc., 96, 299 (1974).
10. W.T. Wipke, G. Grethe, and W. Sieber, Annual report for 1975 to NIH on RR00578.
11. E.J. Corey, W.J. Howe, and D.A. Pensak, J. Amer. Chem. Soc., 96, 7724 (1974).
12. E.J. Corey and W.L. Jorgensen, J. Amer. Chem. Soc., 98, 203 (1976).
13. E.J. Corey, W.J. Howe, H.W. Orf, D.A. Pensak, and G. Petersson, J. Amer. Chem. Soc., 97, 6116 (1975).
14. E.J. Corey and W.L. Jorgensen, J. Amer. Chem. Soc., 98, 189 (1976).
15. E.J. Corey, H.W. Orf, and D.A. Pensak, J. Amer. Chem. Soc., 98, 210 (1976).
16. H. Gelernter, N.S. Sridharan, A.J. Hart, S.C. Yen, F.W. Fowler, and H.J. Shue, Fortschr. Chem. Forsch., 41, 113 (1973).
17. I. Ugi, Rec. Chem. Progr., 30, 289 (1969); Angew. Chem. Internat. Edit., 9, 703 (1970); ibid, 10, 914, 915 (1971); Trans. N.Y. Acad. Sci., 34, 416 (1972).
18. a) J. Degundji and I. Ugi, Fortschr. Chem. Forsch., 39, 19 (1973);
   b) J. Gasteiger, P.D. Gillespie, D. Marguardin, and I. Ugi, Fortschr. Chem. Forsch., 48, 1 (1974);
   c) J. Blair, J. Gasteiger, C. Gillespie, P.D. Gillespie, and I. Ugi, Tetrahedron, 30, 1845 (1974).
19. J. Blair, J. Gasteiger, C. Gillespie, P.D. Gillespie, and I. Ugi in "Computer Representation and Manipulation of Chemical Information," ed. W.T. Wipke, S. Heller, R. Feldmann, and E. Hyde, Wiley New York, N.Y. (1974) p 129.
20. To be presented at New York National American Chemical Society Symposium on "Computer-Assisted Organic Synthesis Planning," April, 1976.
21. J.B. Hendrickson, J. Amer. Chem. Soc., 93, 6847, 6855 (1971).
22. a) J.B. Hendrickson, J. Amer. Chem. Soc., 97, 5784 (1975);
   b) ibid, p. 5763.
23. O. Sinanoglu, J. Amer. Chem. Soc., 97, 2309 (1975).
24. R.V. Stevens and T. Brownscombe, personal communication.
25. J.D. Dubois and H. Viellard, Bull. Soc. Chim. France., 913 (1968).
26. D.H. Smith, B.G. Buchanan, R.S. Engelmore, A.M. Duffield, A. Yeo, E.A. Feigenbaum, J. Lederberg, and Carl Djerassi, J. Amer. Chem. Soc., 94, 5962 (1972).
27. a) R.E. Carhart, D.H. Smith, H. Brown, and N.S. Sridharan, J. Chem. Info. Comp. Sci., 15, 124 (1975);

27. b) R.E. Carhart, D.H. Smith, H. Brown, and C. Djerassi, J. Amer. Chem. Soc., 97, 5755 (1975).
28. J. Schecter and P.C. Jurs, Appl. Spectr., 27, 30 (1973).
29. R.J. Feldmann, S.R. Heller, and K.P. Shapiro, J. Chem. Doc., 12, 41, 48 (1972).
30. a) O. Schier, W. Nubling, W. Steidle, and J. Valls, Angew. Chem. Internat. Edit., 9, 599 (1970); V.R. Fugmann and W. Bitterlich, Chemiker-Zeitung, 96, 323 (1972); V.K-H. Bork, ibid, 96, 330 (1972);
b) M.F. Lynch,"Development and Assessment of an Automatic System for Analysing Chemical Reactions," Final report to the British Library Research and Development Department July, 1975.
31. W.T. Wipke, S.R. Heller, R.J. Feldmann, and E. Hyde, "Computer Representation and Manipulation of Chemical Information," John Wiley, New York, N.Y., 1974.
32. W.T. Wipke, T.M. Dyott, "General Interactive Graphics Language," Princeton Computer Graphics Laboratory, Memo 20, 1-1, April 1971.
33. W.T. Wipke and A. Whetstone, Computer Graphics, 5, 10-22 (1971).
34. W.T. Wipke, C. Still, T.M. Dyott, and P. Friedland, "ALCHEM: A General Purpose Language for Describing Chemical Reactions," in preparation.
35. T. Gund, P.v.R. Schleyer, P.H. G nd, and W.T. Wipke, J. Amer. Chem. Soc., 97, 743 (1975).
36. P. Gund, W.T. Wipke, and R. Langridge, Computers in Chemical Research and Education, Elsever, Amsterdam, Vol. II (1973) pp 5/33-38.
37. For review see J.E. Williams, P.J. Stang, and P.v.R. Schleyer, Ann. Rev. Phys. Chem., 19, 531 (1968).
38. R.G. Powell, D. Weisleder, C.R. Smith, Jr., and I.A. Woeff, Tetrahedron Letters, 4081 (1969); for x-ray see D.J. Abraham, R.D. Rosenstein, and E.L. McGandy, Tetrahedron Letters, 4085 (1969).
39. L.J. Dolby, S.J. Nelson, and D. Senkovich, J. Org. Chem., 37, 3691 (1972).
40. J. Auerbach and S.M. Weinreb, J. Amer. Chem. Soc., 94, 7172 (1972); M.F. Semmelhack, B.P. Chong, and L.D. Jones, ibid, 94, 8629 (1972).

STATEMENT FROM DR. J. LEDERBERG
PROFESSOR OF GENETICS AND
PRINCIPAL INVESTIGATOR: SUMEX-AIM COMPUTER RESOURCE

THIS NOTE IS IN RESPONSE TO THE OUTLINE OF THE OBJECTIVES OF DR. WIPKE'S PROPOSAL FOR A RESOURCE-RELATED RESEARCH PROJECT THAT APPEARS ON PAGE. 25. I HAVE ENJOYED TELEPHONE CONVERSATIONS WITH DR. WIPKE ON HIS PROPOSAL; BUT DID NOT HAVE THE ENTIRE TEXT AT HAND AT THE PRESENT TIME; I WILL BE HAPPY TO RESPOND IN MORE DETAIL TO HIS FORMAL PROPOSAL IF THERE ARE ANY FURTHER QUESTIONS ABOUT ITS RELATIONSHIP TO THE SUMEX AND DENDRAL PROJECTS BASED AT STANFORD.

DR. WIPKE'S WORK HAS BEEN ADJUDGED BY OUR ADVISORY COMMITTEE TO BE ONE OF THE MORE CREATIVE PROJECTS THAT BELONG TO THE SUMEX-AIM USER COMMUNITY. WE WELCOME THE AUGMENTATION OF THIS EFFORT THAT WOULD BE ENABLED BY THE SUCCESS OF HIS APPLICATION. SINCE HE MOVED TO UC/SANTA CRUZ AND BEGIN HIS WORK WITHIN THE SUMEX-AIM FRAMEWORK HE HAS ALSO DISTINGUISHED HIMSELF BY THE COOPERATIVE SPIRIT OF HIS EFFORTS AND HAS BEEN QUITE SUCCESSFUL IN RELATING TO THE CIRCLE OF INVESTIGATORS WORKING ON COMPLEMENTARY PROBLEMS HERE AT STANFORD. HIS ATTRIBUTIONS ABOUT THE COMPLEMENTARY ASPECTS OF HIS WORK; AND OF THE DENDRAL PROJECT; ARE IN ACCORD WITH OUR OWN PERCEPTIONS. WHILE WE HAVE HAD SOME OPPORTUNITY IN THE PAST TO TAKE ADVANTAGE OF HIS PARTICULAR CONTRIBUTIONS TO THE CALCULATION AND DISPLAY OF MOLECULAR GEOMETRIES; THE COOPERATIVE EFFORT WITHIN THE FRAMEWORK OF THE SUMEX-AIM RESOURCE HAS BEEN MUCH MORE EFFICIENT AND PRODUCTIVE. WE BELIEVE THAT OTHER USERS; E.G.; OF THE CONGEN PROGRAM; AND THE XRAY-STRUCTURAL ANALYSIS; WHO WORK ON SUMEX IN RELATED AREAS WILL CONTINUE TO BENEFIT IN A SIMILAR FASHION.

WE HAVE DISCUSSED THE RELEVANT DETAILS OF COST- AND RESOURCE-ALLOCATION WITH DR. WIPKE; AND BELIEVE WE HAVE AN EFFECTIVE ACCOMODATION AS SPELLED OUT IN HIS BUDGET JUSTIFICATION. SPECIFIC ALIQUOTS OF ACCESS TO THE SUMEX RESOURCE MUST BE RE-EXAMINED ANNUALLY BY OUR ADVISORY COMMITTEE; BUT THE LEVEL OF ENTHUSIASM FOR HIS WORK; TOGETHER WITH ACCUMULATING EVIDENCE OF THE COOPERATIVE STYLE OF HIS CURRENT EFFORTS LEAVE ME WITH NO DOUBT THAT HE WILL CONTINUE TO GET HIGH-PRIORITY APPROVAL WITHIN THE FRAMEWORK OF THE FAIR ALLOCATION OF THE RESOURCE TO ALL QUALIFIED USERS. THERE IS NO MANIFEST REASON FOR DR. WIPKE NOT TO RELY UPON SUMEX-AIM AS THE MEDIUM FOR HIS INVESTIGATIONS.

IN HIS OWN BUDGET HE HAS INDICATED THE AREAS WHERE INCREMENTAL CONTRIBUTIONS TO THE OPERATION OF SUMEX-AIM WOULD RELIEVE SERIOUS PRESSURES ON THE RESOURCE FOR THE SPECIFIC BENEFIT OF HIS OWN PROJECT.

THAT WAS MESSAGE 42

<-

DO NOT WRITE IN THIS SPACE-BINDING MARGIN