Current Funding

Mycin is currently in the last year of a three year grant, (HS-01544, Dr. Stanley Cohen, principal investigator) from the Bureau of Health Sciences Research and Evaluation. The grant is for $149,982, and expires May 30, 1977.

Applications pending

A two year renewal of HS-01544 has been submitted to begin June 1, 1977, for $140,000 (direct costs) for the first year. A site visit has been held and the proposal approved but a decision for funding is still pending.

A grant from NSF (Dr. Bruce Buchanan, PI) has been approved for two years, to begin June 1, 1977, for $50,000 a year (direct costs).

A joint application (with Dr. Jon Heiser of UC Irvine) is currently pending with the Biomedical Engineering Division of NIH. The Stanford part of the grant (Dr. Bruce Buchanan, PI) requests a total of $146,751 over 3 years ($46,609 in the first year), to begin June 1, 1977. Dr. Heiser's budget requests $147,655 over 3 years ($46,423 in the first year), to begin July 1, 1977.

A 5-year proposal to the Biotechnology Resources Program is being prepared for submission by June 1, 1977.

II) Interactions with Sumex-Aim resource

Collaborations and medical use of programs

  Dr. Jon Heiser

We have been working with Dr. Jon Heiser of the Department of Psychiatry of the University of California at Irvine, in an effort to create a consultant for the use of psychoactive drugs. We began by creating a version of Mycin that had all of the infectious disease knowledge removed from it, and showed Dr. Heiser how to build up the required base of knowledge about the new field.

He has, with his students, developed a small, but functional system that demonstrates encouraging performance on the task. Work has now begun in earnest to extend the competence of this pilot system, to produce a consultant with a useful level of performance.

It is interesting to note that the explanation capabilities required no modification whatever, and worked in the new system exactly as designed for the original system, despite the change in domains.

INTERNIST Project

The Sumex computer has made possible a valuable interaction between researchers on the MYCIN project at Stanford University and those working on the INTERNIST project at the University of Pittsburgh. These researchers are studying the possible representations and uses for disease models in a medical diagnosis system. Both research groups have been able to run each others programs and to study the medical knowledge bases which are stored on the Sumex computer. Communication between project members has also been greatly facilitated through use of the Sumex system.

Stanford Infectious Disease Faculty

Dr. Victor Yu of our group has been actively soliciting the involvement of the Stanford ID faculty in the development and evaluation of Mycin. He recently presented the system to the faculty and fellows of the Department, and has been seeking ways to involve the system in the Department's educational activities. For instance, medical students under his supervision have used the system during their ID rotation, comparing its results and reasoning process with their own on problems encountered in patients on the wards.

The Pulmonary Function Facility

Members of the Mycin project have also been collaborating with Dr. John Osborn and his co-workers of the Presbyterian Hospital/Pacific Medical Center in San Francisco on the development of a program to interpret the results of standard pulmonary function tests. The program is designed to perform a range of tasks, including: identifying the need to repeat tests because of poor patient effort; identifying the need for additional information in order to make a more definitive diagnosis; reporting and explaining the reasons for primary and secondary diagnoses and severity of any disease state; identifying the relation between diagnosis and any referral diagnosis; and interpreting any change from previous tests, or limitations on the interpretation because of the test methodology and the patient effort.

Sharing with other projects

Groups at Rutgers University, the University of Pittsburgh, Rochester University, and the University of Virginia Medical School have all been involved in varying degrees with running Mycin and evaluating its performance. They have suggested to us improvements in its design, and stock of medical knowledge, and made useful contributions to its development.

In addition, we have made use of the programs developed at both Rutgers and Pittsburgh. The former has been instructive to us in its handling of dynamically changing situations, while the latter has helped us to develop our own ideas about the modelling and use of prototypical descriptions of disease states.

The Molgen group at Stanford has also profited from much of our experience in acquiring knowledge and building large knowledge bases. Several of their

techniques for accumulating knowledge about genetics are based on extensions to ideas first suggested in some of our work.

In all of these cases, the use of Sumex as a national resource has clearly been a critical factor in making possible this sort of interaction.

Critique of resource services

Local management of the existing resources has been carried out in exemplary fashion. The utility of the facilities has consistently increased, as a direct result of the staff's efforts to identify and respond to needs of the user community. They have actively sought out user comments on current and future services and developed programs to support the research work of the community. In particular, the numerous programs for file editing, searching, manipulation, and storage allocation have helped both in data and program management, and in making the best use of available disk storage.

There are, however, additions to the existing resources that would help overcome shortcomings in the available services. In particular, we feel that the addition of more main memory to the system would be an important investment with a significant payoff. First, with the increasing size of the user community, the typical daytime load on the system has increased to the point where running anything but the smallest program requires substantial patience. Second, our project, like several others, is LISP-based, and uses a large address space. Such programs receive lower priority from the scheduler, and especially with the recently changed scheduling algorithm, our effective service level has decreased significantly. The addition of more main memory would ease both of these problems considerably for a number of users.

The addition of more disk space would also be an important improvement in the existing facilities. While it is typically true that disk usage can expand to meet the storage available, we feel that once again the growth of the user community has put a strain on the available resources. We have made extensive use of the archiving facilities, and feel that additional disk space would contribute to the system's utility.

As noted a moment ago, the recently revised scheduling algorithm has also made its impact felt. We have seen our effective service level on the system decrease, as compared to the amount of service we had been getting at a given load average. While we recognize the national scope of the Sumex charter, and the importance of providing adequate service to the whole community, there are a number of major projects located at Stanford. The majority of large projects are thus competing for the same share of the system. It seems unreasonable for, say, three sizable LISP programs to be competing for the same part of the machine, just because they are at Stanford, while a single remote user is receiving nearly all the remaining resource. We recognize the desirability of keeping Sumex a national resource, but wonder if there is a way this can be done without penalizing systems just because they originate at Stanford.

Finally, there is a smaller scale project which would also make a substantive contribution to the utility of the resource. Currently a program called PUB is the major text formatting ("word processing") program in use. It

is something of an historical relic, and is quite large, not totally reliable, and rather difficult to use. It is remarkably powerful, but most users make relatively little use of its more impressive powers. Since preparing technical reports, progress reports, and thought-pieces on proposed or in-progress work are all an integral part of doing research, facilities that ease the task can make an important contribution to the progress of work. A new program, designed along the lines of PUB, but much smaller and of proven reliability, would be an important contribution to the research efforts of the community. It would require on the order of one man-year to create, but given the anticipated drain on system resources presented by the amount of technical writing done by the community, this investment would quickly be paid back many times over.

III)   Follow-on

Long range project goals

        The long-term goals of our project center around further development of our ideas on computer-based medical consultants. We intend, for instance, to extend both the depth and breadth of the system's range of competence. The extension in breadth will be an important demonstration of the power of the approach we have taken, since the problem of scale is a traditional pitfall that has trapped a number of other efforts in AI. We believe that our techniques provide the basis for continued effective performance, even with a much larger knowledge base that handles a wider scope of medical problems. This can only be tested, of course, by actually enlarging the knowledge base and widening the program's scope.

        By extending the "depth" of the program's competence, we mean dissecting still further the concepts on which its judgments are based. The current system, for instance, asks the doctor if the patient is "febrile due to the infection". In practice, this is a difficult judgment to make, and it is precisely on such difficult judgmental issues that Mycin should be able to offer assistance. By asking our clinicians to specify how they decide that a patient is febrile due to an infection, we can break down this vague notion into a number of distinct decision rules. The resulting program will make fewer demands on the user, and hence will offer a more effective source of consultative advice.

        We also believe that the best hardware for many AI research efforts lies in the direction of independent minicomputers arranged as a satellite to a central system, and capable of running high level languages (like LISP). A second of our long-term goals, then, is to develop a version of our program capable of running on such a system. Since there are currently a number of efforts aimed at developing both high level languages for mini-machines, and minicomputer architectures capable of running high level languages, Sumex could benefit substantially from this work if the AIM Committee begins now to plan to take advantage of these developments.

        We also plan to extend the generality of the system we have developed, to make it possible for experts in other medical (and medically-related) areas to

use it as a framework for assembling their own set of decision rules, to create
consultants for their own specialties. We have already attempted several pilot
studies along these lines (the work with Dr. Heiser on psychopharmacology, and
with Dr. Osborn on pulmonary function). Each of these has demonstrated to us a
number of generalizations that our current techniques require. We plan to make
these changes, and continue to develop a system usable by a wide range of
specialists, as part of our interest in the art of building expert systems.

A necessary parallel development to this will be improvements in the rule-
based representation of knowledge and a better understanding of the process of
clinical decision making. While our decision rules offer a number of advantages,
we have also seen some drawbacks in them, and plan to work on overcoming the
problems without losing the advantages they offer. Our present model of decision
making under uncertainty is still elementary and intuitive -- further work is
needed to make it more formal and ground it firmly in well understood principles.
This will also facilitate work on other problem, such as checking the internal
consistency of the entire set of rules.

Justification

Our project is concerned with a range of problems that are central to both
medical care and AI research. Earlier sections of this report covered the
significance of the specific problem of antibiotic misuse. More generally, the
problem of medical decision making is one that has received much attention, and
has not yet yielded to a definitive solution. The availability of computer-based
advisors for difficult clinical problems would be a useful step in combatting the
current maldistribution of specialists. With network links to centralized
machines, or mini-machines inexpensive enough to be exported as a unit, hospitals
in outlying rural areas might have available a sophisticated source of medical
advice.

The development of computer-based consultants is a mainstream issue in AI
research. Its specific goals are to produce expert performance on a "real world"
problem, and to make that expertise available to users who might not normally be
involved with computers. Producing a system that both offers high performance
and presents a reasonable interface to the user means solving a difficult problem
with a number of constraints. High performance alone is not enough, since the
system must be usable by a computer-naive audience. This means more than simply
reasonable I/O facilities, and implies the need for such things as the
explanatory capabilities currently a part of Mycin.

More generally, the issue of accumulating, representing, and using large
stores of task-specific knowledge is an important thrust of current AI research.
Ever since the failure of the original GPS-type approach to problem solving (in
which problem solving power comes from a single, domain-independent paradigm),
interest has been focussed on the use of large stores of domain-specific
knowledge as a source of high performance. This has been a primary theme of the
work on Mycin from the outset, and our efforts have produced a number of insights
about the design and construction of such systems. We have emphasized, for
instance, the importance of keeping a sharp distinction between the base of task-
specific knowledge and the interpreter which uses that information to solve
problems. This design pays off both by easing the task of building the knowledge

base, and by increasing the range of applicability of the underlying system (i.e., different knowledge bases can be "plugged in" to the same underlying system).

Finally, a number of other projects have been "spun off" as a direct result of ours. The pulmonary function work and the work by Dr. Heiser's group are both outgrowths of Mycin, and have both begun to produce their own substantive results.

Future resource goals

As noted earlier, we see the development of minicomputers that run high level languages as an important future trend that will affect much of the work in AI. We believe it will be especially advantageous for Sumex to take advantage of these developments. Adding a small number of these minicomputers as satellites to the main system would present a number of important advances. First, many of the research efforts currently underway involve large, LISP-based programs that significantly impact the system load. By providing satellite machines to which those large systems could be shifted, the system load would lighten considerably and the large systems would themselves run much faster. Second, it would mean more efficient use of resources, since adding these satellite systems would require little or no additional tapes, disks, printers, etc. Finally, many projects are in a situation parallel to ours, in that work proceeds on two fronts simultaneously. One one hand, new ideas are being generated about how a program should work, or what tasks it might perform. These are implemented and tried out in a test version of the program. On the other hand, once those ideas prove practical, there is often an extensive period of development that requires a more stable version of the program. The architecture suggested here, of a main system with satellite machines, offers an excellent environment for this work, since smaller test versions of a program can be used as a "proving ground" on the main machine, while the larger, stabilized versions are further developed by running them on the satellite machines.

The sort of arrangement is most effective when transition between systems is almost invisible -- that is, when little or nothing need be done to shift from the central machine to a satellite. This is easiest to do when there are high-bandwidth data links betwen machines, and satellite machines capable of running the same programming language as the central machine.

We believe it would be important to provide Sumex support for both the software as well as the hardware problems involved in creating this sort of environment. One effort in this direction (Mainsail) is currently underway, and parallel efforts at other locations are involved in producing a version of LISP that will run on small machines. While there is no need to duplicate these latter efforts, we feel it would be important for Sumex to stay closely coupled to them, so that their results can easily and quickly be implemented here. Given the number of projects which could make significant use of these results, and the impact those projects currently have on the system, we believe the investment in time and effort would pay off quite well.

References

[1] Reiman H H, D´ambola J, The use and cost of antimicrobials in hospitals, Arch
    Environ Health, 13:631-636 (1966).

[2] Kunin C M, et.al., Use of antibiotics: a brief exposition of the problem and
    some tentative solutions, Anns Int Med, 79:555-560 (1973).

[3] Sheckler W E, Bennett J V, Antibiotic usgae in seven community hospitals, J
    Amer Med Assoc, 213:264-267 (1970).

[4] Roberts A W, Visconti J A, The rational and irrational use of systemic
    antimicrobial drugs, Amer J Hosp Pharm, 29:828-834 (1972).

[5] Simmons H E, Stolley P D, This is medical progress? Trends and consequences
    of antibiotic use in the United States, J Amer Med Assoc, 227:1023-1026
    (1974).

[6] Kagan B M, Fanin S L, Bardie F, Spotlight on antimicrobial agents, JAMA,
    226:306-310 (1973).

6.1.5   PROTEIN STRUCTURE PROJECT

Protein Structure Modeling Project

Prof. J. Kraut and Dr. S. Freer (Chemistry, U. C. San Diego) and
Prof. E. Feigenbaum and Dr. R. Engelmore (Computer Science, Stanford)

I. Summary of research program

A. Technical goals

The goals of the protein structure modeling project are to 1) identify
critical tasks in protein structure elucidation which may benefit by the
application of AI problem-solving techniques, and 2) design and implement
programs to perform those tasks. We have identified two principal areas which
have both practical and theoretical interest to both protein crystallographers
and computer scientists working in AI. The first is the problem of interpreting
a three-dimensional electron density map. The second is the problem of
determining a plausible structure in the absence of phase information normally
inferred from experimental isomorphous replacement data. Current emphasis is on
the implementation of a program for interpreting electron density (e.d.) maps.

B. Medical relevance and collaboration

The biomedical relevance of protein crystallography has been well stated in
a recent textbook on the subject (Blundell & Johnson, Protein Crystallography,
Academic Press, 1976):

> "Protein Crystallography is the application of the
> techniques of X-ray diffraction ... to crystals of one
> of the most important classes of biological molecules,
> the proteins. ... It is known that the diverse
> biological functions of these complex molecules are
> determined by and are dependent upon their
> three-dimensional structure and upon the ability of
> these structures to respond to other molecules by
> changes in shape. At the present time X-ray analysis of
> protein crystals forms the only method by which detailed
> structural information (in terms of the spatial
> coordinates of the atoms) may be obtained. The results
> of these analyses have provided firm structural evidence
> which, together with biochemical and chemical studies,
> immediately suggests proposals concerning the molecular
> basis of biological activity."

The project is a collaboration of computer scientists at Stanford
University and crystallographers at the University of California at San Diego
(under the direction of Prof. Joseph Kraut) and at Oak Ridge National
Laboratories (Dr. Carroll Johnson).

C. Progress summary

During the past year we have been designing and implementing a system of programs for interpreting three-dimensional e.d. maps. Progress has been made by attacking the problem from two directions: working upward from the primary data (i.e. the array of e.d. values) to higher level symbolic abstractions, and working downward from the given amino acid sequence and other experimental information to generate candidate structures which can then be confirmed by the abstracted data.

In the "bottom-up" area of research we have developed and implemented programs for analyzing topological features of the skeletonized e.d. map in terms of protein structural elements (e.g., side chains, chain ends, bridges, etc.), for finding local maxima, and, recently for generating a critical point network, i.e. a three-dimensional spanning tree which connects all critical points (peaks, saddle points) found in the map.

In the "top-down" area we have designed and implemented, in INTERLISP, a structure inference program which generates structural hypotheses at several levels of detail. At present the program can infer, from the amino acid sequence and other chemical information, and the symbolic abstractions of the e.d. map, the location of heavy atoms, cofactors and chain ends. Those features provide toeholds, i.e. islands of certainty, from which additional structure is inferred by extension. Work is currently in progress on identification of the main chain, disambiguation of multiply connected regions and classification of side chain regions.

The system under development is knowledge-based. Both the corpus of knowledge of the task domain and the problem-solving strategy knowledge are incorporated as production-like rules.

D. List of Publications

1) Robert S. Engelmore and H. Penny Nii, "A Knowledge-Based System for the Interpretation of Protein X-Ray Crystallographic Data," Heuristic Programming Project Memo HPP-77-2, January, 1977. (Alternate identification: STAN-CS-77-589)

2) E.A. Feigenbaum, R.S. Engelmore, C.K. Johnson, "A Correlation Between Crystallographic Computing and Artificial Intelligence," in Acta Crystallographica, A33:13, (1977). (Alternate identification: HPP-77-25)

E. Funding status

The project recently received a renewal of its funding from the National Science Foundation. The new research period began on May 1, 1977, and is for a two year period at a funding level of $75,000 per year. No other applications are pending.

II. Interaction with the SUMEX-AIM resource

A. Collaborations

The protein structure modeling project has been a collaborative effort since its inception, involving co-workers at Stanford and UCSD (and, more recently, at Oak Ridge). The SUMEX facility has provided a focus for the communication of knowledge, programs and data. Without the special facilities provided by SUMEX the research would be seriously impeded. Computer networking has been especially effective in facilitating the transfer of information. For example, the more traditional computational analyses of the UCSD crystallographic data are made at the CDC 7600 facility at Berkeley. As the processed data, specifically the e.d maps and their Fourier transforms, become available, they are transferred to SUMEX via the FTP facility of the ARPA net, with a minimum of fuss. (Unfortunately, other methods of data transfer are often necessary as well -- see below.) Programs developed at SUMEX, or transferred to SUMEX from other laboratories, are shared directly among the collaborators. Indeed, with some of the programs which have originated at UCSD and elsewhere, our off-campus collaborators frequently find it easier to use the SUMEX versions because of the interactive computing environment and ease of access. Advice, progress reports, new ideas, general information, etc. are communicated via the message and/or bulletin board facilities.

B. Interaction with other SUMEX-AIM projects

Our interactions with other SUMEX-AIM projects have been mostly in the form of personal contacts. We have strong ties to the DENDRAL, Meta-DENDRAL and MOLGEN projects and keep abreast of research in those areas on a regular basis through informal discussions. The SUMEX-AIM workshop in June, 1976 provided an excellent opportunity to survey all the projects in the community. Common research themes, e.g. knowledge-based systems, as well as alternate problem-solving methodologies were particularly valuable to share. (That workshop was very likely the most significant conference for applied AI to be held in 1976.)

C. Critique of Resource services

On the whole the services provided by SUMEX have been excellent, considering the large demand on its resources. With the important exceptions of high peaks in the weekday prime-time load average, the ratio of CPU time to total wait time during program execution is usually acceptable. The facility provides a wide spectrum of computing services which are genuinely useful to our project -- message handling, file management, Interlisp, Fortran and text editors come immediately to mind. Moreover, the staff, particularly the operators, are to be commended for their willingness to help solve special problems (e.g., reading tapes) or providing extra service (e.g., and immediate retrieval of an archived file). Such cooperative behavior is rare in computer centers.

A serious fault in the system is the lack of reliable tape drives, and the paucity of the present software for handling tape files. Much of our data from the outside world is received on magnetic tape, and almost never in the unusual

PDP-10 format. We urge that the existing tape drives be replaced, and software be provided to facilitate the input of data in non-standard formats. (At the present time there is not even a program to provide a byte-by-byte dump when all else fails.)

III. Use of SUMEX during the follow-on grant period (8/78 - 7/83)

   A. Long-range goals

        Our current research grant extends through April, 1979. During that time we intend to bring the structure modeling system to a level of performance that permits reliable qualitative interpretation of high resolution e.d. maps, derived from real data and a correct amino acid sequence. We also plan to exploit the flexibility of the rule-based control structure to permit investigation of alternate problem-solving strategies and modes of explanation of the program's reasoning steps. Beyond the next two years, emphasis will be placed on expanding and generalizing the system to relax the constraints of resolution and accuracy in the input data.

   B. Justification for continued use of SUMEX

        The biomedical relevance of the protein structure modeling project, coupled with the need for building a computational system with a significant component of symbolic inference, qualifies the project as an AIM-relevant endeavor. SUMEX provides an excellent computing environment for creating and debugging programs (in a variety of languages), for sharing and distributing information among geographically dispersed co-workers, and for keeping up with current research in other AIM areas. Our project is clearly too small to justify an independent computing facility, and other large computer centers that are conveniently accessible do not fulfill our requisites. Consequently SUMEX has been and hopefully will continue to be an integral research tool in this project.

   C. Comments and suggestions

        Two improvements to the system which, though not critical, would appreciably upgrade the service provided:

1. Connection of SUMEX to a non-military network which permits file transfer at a reasonably high rate (at least 4800 baud). The restrictions imposed on the use of the ARPA network prohibit using it to transmit large program and/or data files between SUMEX and the UCSD computing facilities. The availability of such a connection would, for example, permit us to use their E&S interactive graphics system to display and visually examine the structures hypothesized by our automated modeling system.

2. Addition of 256K of main memory, to give more rapid response during the peak hours. This would seem to be a natural extension to the system, to complement the second KI-10 installed last year, and would more fully realize the potential of the second CPU.

6.2   <u>NATIONAL</u> <u>AIM</u> <u>PROJECTS</u>

The following group of projects is formally approved for access to the AIM
aliquot of the SUMEX-AIM resource. Their access is based on review by the AIM
Advisory Group and approval by the AIM Executive Committee.

6.2.1   ACQUISITION OF COGNITIVE PROCEDURES (ACT)

Acquisition of Cognitive Procedures (ACT)

Dr. John Anderson
Yale University

(Grant NIMH MH29353 $25,000 this year)
(Contract ONR N0014-77-6-0242 $74,000 this year)

I. Summary of Research Program

A. Technical goals:

To develop a production system that will serve as an interpreter of the
active portion of an associative network. To model a range of cognitive tasks
including memory tasks, inferential reasoning, language processing, and problem
solving. To develop an induction system capable of acquiring cognitive
procedures with a special emphasis on language acquisition.

B. Medical relevance and collaboration:

1. The ACT model is a general model of cognition. It provides a useful
    model of the development of and performance of the sorts of decision
    making that occur in medicine.

2. The ACT model also represents basic work in AI. It is in part an attempt
    to develop a self-organizing intelligent system. As such it is relevant
    to the goal of development of intelligent artificial aids in medicine.

We have been evolving a collaborative relationship with Dr. James Greeno
and Allan Lesgold at the University of Pittsburgh. They are applying ACT to
modeling the acquisition of reading and problem solving skills. We plan to make
ACT a guest system within SUMEX. ACT is currently at the state where it can be
shipped to other INTERLISP facilities. We have received a number of inquiries
about the ACT system. ACT is a system in a continual state of development but we
periodically freeze versions of ACT which we maintain and make available to the
national AI community.

C. Progress and accomplishments:

ACT provides a uniform set of theoretical mechanisms to model such aspects
of human cognition as memory, inferential processes, language processing, and
problem solving. ACT's knowledge base consists of two components, a
propositional component and a procedural component. The propositional component
is provided by an associative network encoding a set of facts known about the
world. This provides the system's semantic memory. The procedural component

consists of a set of productions which operate on the associative network. ACT's
production system is considerably different than many of the other currently
available systems (e.g., Newell's PSG). These differences have been introduced
in order to create a system that will operate on an associative network and in
order to accurately model certain aspects of human cognition.

A small portion of the semantic network is active at any point in time.
Productions can only inspect that portion of the network which is active at the
particular time. This restriction to the active portion of the network provides
a means to focus the ACT system in a large data base of facts. Activation can
spread down network paths from active nodes to activate new nodes and links. To
prevent activation from growing continuously there is a dampening process which
periodically deactivates all but a select few nodes. The condition of a
production specifies that certain features be true of the active portion of the
network. The action of a production specifies that certain changes be made to
the network. Each production can be conceived of as an independent "demon." Its
purpose is to see if the network configuration specified in its condition is
satisfied in the active portion. If it is, the production will execute and cause
changes to memory. In so doing it can allow or disallow other productions which
are looking for their conditions to be satisfied. Both the spread of activation
and the selection of productions are parallel processes whose rates are
controlled by "strengths" of network links and individual productions. An
important aspect of this parallelism is that it is possible for multiple
productions to be applied in a cycle through the set of productions. Much of the
early work on the ACT system was focused on developing computational devices to
reflect the operation of parallel, strength-controlled processes and working out
the logic for creating functioning systems in such a computational medium.

We have successfully implemented a number of small-scale systems that model
various psychological tasks in the domain of memory, language processing, and
inferential reasoning. A larger scale effort is underway to model the language
processing mechanisms of a young child. This includes implementation of a
production system to analyze linguistic input, make inferences, ask and answer
questions, etc. Also a great deal of effort is being given to developing
learning mechanisms that will acquire and organize the productions for this
language processing. This learning program attempts to acquire procedures from
examples of the computations desired of the procedures. For instance, the
program learns to comprehend and generate sentences by being given sentences and
picture representations of the meaning of the sentences (actually hand encodings
of the pictures). Although this effort is focused on induction of linguistic
procedures, the hope is to develop a general model of induction of cognitive
procedures and not to place any language-specificity into the induction
procedures.

At the time of this report, we have completed the F version of ACT which is
the system with learning capabilities. We are currently testing and tuning the
system on a number of linguistic examples. Other projects which are progressing
in earlier versions of ACT include use of spreading activation to model semantic
disambiguation, modeling of the reading process, and modeling of solutions to
word arithmetic problems.

D. Current list of project publications:

[1] Anderson, J.R. Computer simulation of a language acquisition system: A
    second report. In D. LaBerge and S.J. Samuels (Eds.). Perception and
    Comprehension. Hillsdale, N.J.: L. Erlbaum Assoc., 1975.

[2] Anderson, J.R. Language, Memory, and Thought. Hillsdale, N.J.: L. Erlbaum,
    Assoc., 1976.

[3] Anderson, J.R. Induction of augmented transition networks. Cognitive
    Science, 1977, in press.

[4] Anderson, J.R. & Kline, P. Design of a production system. Paper to be
    presented at the Workshop on Pattern-Directed Inference Systems, Hawaii, May
    23-27, 1977.

[5] Anderson, J.R., Kline, P. & Lewis, C. Language processing by production
    systems. To appear in P. Carpenter and M. Just (Eds.). Cognitive Processes
    in Comprehension. L. Erlbaum Assoc., 1977.

[6] Kline, P.J. & Anderson, J.P. The ACTE User's Manual, 1976.

II. Interaction With the SUMEX-AIM Resource

The SUMEX-AIM resource is superbly suited for the needs of our project. We
have made the most extensive use of the INTERLISP facilities and the facilities
for communication on the ARPANET. We have found the SUMEX personnel extremely
helpful both in terms of responding to our immediate emergencies and in providing
advice helpful to the long-range progress of the project. Despite the fact that
we are on the other side of the continent, we have felt almost no degradation in
our ability to do research. We find we can easily list on the terminal a small
portion of programs under modification. The willingness of SUMEX mail listing
has also meant we can keep relatively up-to-date records of all programs under
development.

A unique east coast advantage of working with SUMEX is the low loading of
the system during the mornings. We have been able to get a great deal of work
done during these hours and try to save our computer-intensive work for these
hours.

We have found our one AIM work shop so far (1976) a very useful opportunity
to meet with colleagues and exchange ideas.

A particularly striking example of the utility of the SUMEX resource was
illustrated in the move from Michigan. In the summer of 1976 Anderson moved to
Yale and Greeno to Pittsburgh. There was no loss at all associated with having
to transfer programs from one system to another. At Yale we were programming the
day after we arrived. The SUMEX link has also permitted continued collaboration
with Greeno.

From our point of view, the only stress in the SUMEX resource involves the
issue of the tight file space. We are managing with some difficulty to stay

within our allocation. We do not feel that our allocation is unfair (in light of overall availability and number of users) but we do feel that all users would be able to work more comfortably with the system if there were more file space available. While we recognize the need for purges when projects exceed their allocation, we feel that it would be useful if this purge could be made more intelligent, perhaps purging according to a user defined priority.

III. Follow-On SUMEX Grant Period (8/78-7/83)

   A. Long-range user project goals and plans:

       Our long-range goals are: (1) continued development of the ACT system; (2) application of the system to modeling of various cognitive processes; (3) Dissemination of the ACT system to the national AI community.

1. System Development: We are just completing the F version of the ACT system. We fully anticipate that its design will undergo considerable change after we have explored and tested its empirical consequences. We are currently applying or intend to apply the ACT system to modeling the acquisition and/or performance of cognitive skills in the areas of language comprehension and generation, inferential reasoning, reading skills, problem solving, and memory retrieval. It is hard to anticipate now the impact of these explorations for design decisions in later versions of ACT. However, it is clear now that we will want to make ACT more appropriate as a language for programming cognitive skills. This will involve such things as development of more powerful control conventions, simplication of syntax, and introduction of direct programming features (such as comparison of quantity magnitudes) that can only be obtained indirectly now. We would also like to introduce more efficient implementation techniques to replace some of the simple devices that were used to enable us to rapidly complete the system. These rearchitecture efforts have to be done within the constraints of psychological plausibility, but we have a theoretical commitment to the conjecture that good implementation design is predictive of good psychological mechanisms.

2. Application of Modeling Cognitive Processes: We anticipate a gradual decrease in the amount of effort that will go into system development and an increase in the amount of effort that will go into application of the system for modeling. We mentioned above the modeling efforts that we are using to assess the suitability of the ACTF system. We have long-range commitments to apply the ACT learning model to the following three topics: Acquisition of language (both first and second language acquisition); acquisition of reasoning and memory-management skills for geography; acquisition of problem solving skills in the domain of geometry. We find each of these topics to be considerable interest in and of themselves, but they also will serve as strong tests of the learning model. We are hopeful that the systems that are acquired by ACT will satisfy computational standards of good artificial intelligence. Therefore, in future years we would also be interested in applying the ACT model to acquisition of cognitive skills in medically related domains such as diagnosis or scientific inference. SUMEX would be an ideal location for collaboration on such a project.

3. <u>Dissemination of the ACT Project:</u> We have a commitment to making the ACT
   system available to anyone in the national community who has access to the
   necessary computer resources. This is partially to provide a service in that
   ACT is a medium for psychological modeling. However, it is also self-serving
   in that the use of other people make of ACT has important feedback in
   assessing design decisions. In light of limitations on the SUMEX resource,
   we have decided not to allow extensive use of ACT by other researchers
   through our SUMEX account. We feel that extensive use of the ACT system in
   SUMEX by another researcher must have the status of an independent project
   and must be able to justify independently its use of the SUMEX-AIM resource.

B. Justification for continued use of SUMEX:

     We feel that the justification for our use of SUMEX has only been
strengthened since the time of our original application for user status. The
project meets a number of criteria for SUMEX relevance: Project support comes
from NIMH. The project is concerned with cognitive modeling which is a SUMEX
goal. The project is also developing an AI tool which can be used to help
automate various medically-relevant tasks. We also think we are the type of need
that the SUMEX facility was designed to meet. That is, we do not have nearly as
powerful computing facilities local at Yale; we are non-local user; we are using
SUMEX as a base for collaborating with scientists in other parts of the country;
and we are trying to develop a system that will be of general use.

C. Comments and suggestions for future resource goals:

     We would, of course, be delighted if the computational capacity of the
SUMEX facility could be increased. We suffer most severly with the file space
limitation. The other limitation is the slowness of the system at peak hours.
This problem is perhaps less grievous for us than Stanford-based users because of
our ability to use morning hours. We do not feel any urgent need for development
of new software. Our work is growing to such a size that we would find it useful
to have a local ARPANET tip. We are currently discussing this possibility with
our ONR officials. Such a tip might be justifiable given additional needs of
other AI people at Yale. The consequence of such a TIP for the future planning
of SUMEX resources is that we would then change our access to SUMEX from the
TYMNET to the ARPANET, thus relieving SUMEX of the need to support our TYMSHARE
costs.

6.2.2   CHEMICAL SYNTHESIS PROJECT (SECS)

SECS - Simulation and Evaluation of Chemical Synthesis

W. Todd Wipke
Department of Chemistry
University of California at Santa Cruz

I. Summary of Research Program

A. Technical Goals.

The long range goal of this project is to develop the logical principles of
molecular construction and to use these in developing practical computer programs
to assist investigators in designing stereospecific syntheses of complex bio-
organic molecules. Our specific goals this past year focused on improvement of
the library of chemical transforms, completion of the perception of molecular
symmetry and integrating the use of symmetry information throughout SECS
including the strategy module. We also wanted to improve the execution speed of
SECS, and the speed of graphical interaction over remote communication lines. We
planned to simplify the program from the user's viewpoint by including automatic
file failsafing, improvement of HELP commands, and non-fatal handling of all
errors, as well as production of user's manuals for operation of the program and
the writing of chemical transforms. Additionally we intended to initiate
applications of SECS to the areas of biosynthesis and metabolism of compounds, as
well as phosphorus chemistry. Finally we hoped to improve the strategic
constraints and controls that guide SECS in growing a synthesis tree.

B. Medical Relevance and Collaboration.

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
molecules as well as his ability to modify existing structures, e.g.,
incorporating isotopic labels into biomolecular substrates. The Simulation and
Evaluation of Chemical Synthesis (SECS) project aims at assisting the chemist in
designing stereospecific syntheses of biologically important molecules. The
advantages of this computer approach over a manual approaches are manyfold: 1)
greater speed in designing a synthesis; 2) freedom from bias of past experience
and past solutions; 3) thorough consideration of all possible syntheses using a
more extensive library of chemical reactions than any individual person can
remember; 4) greater capability of the computer to deal with the many structures
which result; and 6) capability of computer to see molecules in graph theoretical
sense, free from bias of 2-D projection.

SECS was designed to be able to apply any kind of chemical transformation,
and because of this generality we see SECS finding application in biogenesis and
metabolism (see section II A below). The objective of using SECS in biogenesis
is to predict possible biogenetic pathways for a given natural product and also

to predict related compounds which might also co-occur in nature. This can be a
great aid in searching for new natural products and in structure elucidation.

The objective of using SECS in metabolism is to predict the plausible
metabolites of a given xenobiotic in order that they may be analyzed for possible
carcinogenicity. Metabolism research may also find this useful in the
identification of metabolites in that it suggests what to look for, and in the
identification of possible metabolic pathways connecting a metabolite to a
xenobiotic.

C. Progress and Accomplishments.

RESEARCH ENVIRONMENT: At the University of California, Santa Cruz, we have
a GT-40 graphics terminal connected to the SUMEX-AIM resource by a 1200 baud
leased line and a TI 725 thermal printing teletype connected via TYMNET at 300
baud. UCSC has only a small IBM 370/145 and a PDP-11/45 (limit of 12 K words per
user) available, both of which are unsuitable for this research. From July until
December our research group had to occupy temporary space during renovation, but
is now finally in permanent space in Thimann Laboratories where we have close
collaboration with other organic chemists.

CHEMICAL TRANSFORMS: The library of chemical transforms has been
reorganized and reevaluated during the past year by Mr. Dolata, a student of
Professor D.A. Evans of Cal Tech. New reactions were added and the scope and
limitations of others were updated and leading references provided.
Additionally, Merck, Sharp, and Dohme Research Laboratories provided revisions of
many transforms which a group of 25 synthetic chemists had carefully researched.

SYMMETRY: An efficient algorithm for recognizing molecular symmetry was
developed last year. This year that algorithm has been tested against all
possible molecular point groups and a few problems which developed were
corrected. The algorithm has been documented and initial studies begun on
actually determining the point group of a molecule. The symmetry group is now
utilized in conjunction with the symmetry of a chemical transform so the
transform is applied in all possible unique ways, to generate a non-redundant set
of precursors. This symmetry of course takes into account stereochemistry of
saturated centers and double bonds. We have surveyed literature syntheses for
examples of existing heuristics based on symmetry which can be used for
automatically generating high level strategies. This information has never been
pulled together before and should make an interesting contribution also to
organic synthesis.

STRATEGIC CONTROL: Last year we began developing an implementation of
strategic control for SECS, and a simple language for expressing strategies
independent of chemical transforms. Since these strategies contain expressions
which refer to the molecular structure, it was also necessary to incorporate
symmetry here too. For example, if a particular bond is designated as strategic
to break, but a transform breaks another bond, the strategy is still satisfied if
the two bonds are equivalent by symmetry. This problem becomes more complex when
pairs of bonds are specified and when there are logical connectives (AND, OR,
XOR, and NOT) involved. This has however been solved. Other changes since last
year include a completely new user interface to strategy to allow error

correction and very easy modification of goals. Finally quantitative experiments
have been performed to measure the effect of developing a synthesis tree with
various types of strategic constraints. The net result of this work is that the
user can more easily constrain SECS now to work only in areas which the user
decides are worthwhile, consequently fewer precursors are generated which the
user would delete.

USER INTERFACES: Users of SECS had difficulty understanding how to copy
files into work areas in order to save or restore synthesis trees. Now SECS does
all file manipulation, eliminating the problem. Further SECS now automatically
failsafes the synthesis tree at key points so that in the event of machine or
communication failure the user can automatically restart his analysis from the
last key point. Considerable modifications were made to the graphical interface
for increasing readability and speed of interaction. Over long slow
communication lines (which happens to be the way most SECS users are accessing
the program) interactive graphics must be done with care, minimizing the amount
and frequency of picture transmission, in order to achieve even tolerable man-
machine communication. Lastly, we have implemented appropriate input procedures
to eliminate the possibility of a fatal crash from user input errors. According
to user reports this was a major problem.

PHOSPHORUS CHEMISTRY: Graphical input and output procedures were developed
for entering the stereochemical configuration of a trigonal bipyrimid (TBP)
phosphorus atom and for producing a correct structural diagram from the machine's
internal representation. The SEMA algorithm for generating a stereochemically
unique name was extended to deal with the 20 possible configurations for each TBP
center, including the ability to recognize enantiomers. The ALCHEM language for
representing chemical transforms was extended to facilitate manipulation of
TBP's, including changes from trigonal and tetrahedral configurations to square
base pyramid and TBP. Queries may deal with apicophilicity, and axial or
equatorial orientation. The fine details of phosphorus chemistry such as the
fact that groups entering or leaving the phosphorus coordination sphere normally
do so from the apical position. Pseudo rotation, apicophilicity, and strain
energy are considered in evaluating the stable TBP configurations and in checking
for identical structures. A library of phosphorus chemistry is now being
prepared in collaboration with a group at the University of Strasbourg, France.

COMPUTER-AIDED ELUCIDATION OF BIOGENETIC PATHWAYS: Although a great amount
of effort has been spent on various areas of biogenesis, there have been few
attempts to develop general techniques for the elucidation of biogenetic schemes.
As a result, the formulation of biogenetic schemes has often been criticized for
its lack of rigor and explicit criteria. Our approach is to develop general
techniques which lead to the postulation of plausible biogenetic pathways, using
the SECS as an aide in obtaining and analyzing solutions to this complex problem.
It is our hope this application of computer problem solving techniques will not
only uncover new ways of recognizing and evaluating biogenetic pathways but also
provide added support to deductions made from biogenetic schemes, such as the
generality of a scheme which may be tested in only a few species.

With the proper input information and goals well defined there may be
explicit rules to guide the chemist to plausible biogenetic pathways for a
particular natural product. Unfortunately, the vast majority of solutions to
this problem are determined by a combination of the experienced natural products

chemist's ability to consider the most important rules involved and his unique set of experience-based prejudices. There may be some means to represent and utilize all of the known relevant rules, data and possibly even experience-based prejudices to arrive at the best plausible pathways. The most precise method for representing, developing and testing such a theory is in the form of a computer program. To implement such a computer program, known rules and constraints must be clearly defined, then those that are applicable can be applied at each step of the analysis toward the desired goal. This will keep the solution pathways logically pure and insure that all alternatives which satisfy the rules and constraints are considered. This guarantee of completeness simply can not be made using hand analysis.

A new reaction library containing biogenetic transformations have been written. After inputting a natural product the program will apply the biogenetic transforms which fit the natural product. This generates a set of plausible biogenetic precursors to the target natural product. By continuing this process with the precursors generated, the plausible biogenetic pathways for the natural product can be discovered.

The structures of marine natural products were entered into the program and the plausible biogenetic pathways for these compounds were generated and analysed. Biogenetic pathways which had been proposed in the literature were among the pathways discovered, as were other plausible pathways which would now have to be considered. The success we attained in this research effort verified the applicability of the SECS program as an aid in the analysis of metabolic pathways.

COMPUTER-AIDED PREDICTION OF METABOLITES FOR CARCINOGENICITY STUDIES: We have initiated a research project in collaboration with the Chemical Carcinogenesis group at the National Cancer Institute. The objective of this research is to establish a computer program by which a biochemist or metabolism expert can explore the metabolism of a chemical compound. The investigator enters the substrate molecule by interacting with an input and structure editing module.

Then the program will apply the biological transforms which "fit" the structure, taking into consideration all the context information (2-D, 3-D, and electronic) available about the transform and all perceived information about the structure. This will generate a set of metabolites which are one step away from the substrate structure. The metabolites will be ranked according to expected probability or yield. The exact parameters which should be monitored will be determined during the course of this research. An evaluation module may then screen these metabolites according to criteria specified by the investigator. Duplicate metabolites arising from different pathways will be labelled to indicate that fact. Finally the investigator will be shown the set of metabolites together with data about the transform which produced each one and the values of the parameters being monitored.

The investigator may select one metabolite for further metabolism or may request that all be processed for a specified number of steps. In this way a "tree" of metabolites is produced and displayed. The entire state of the user's tree may be saved to permit continuation of the analysis at another time. Exploration of the metabolism tree will be predominately guided interactively by

the expert investigator. We feel that at this stage of development of the field
of metabolism and carcinogenicity that interactive guidance by the expert is
necessary. There are many areas where the theory is very thin and a given
biological transformation may have been observed for only a few substrates. When
this transform is applied to a new substrate, some unrealistic metabolites may be
generated owing to the deficiency of contextual information and constraints. An
expert is necessary to prune the tree and prevent the automatic processing of
those unreasonable intermediates. It is much more efficient for the expert to do
this pruning as the tree is being grown, rather than later after an enormous tree
has been completed.

At some point either during tree generation or at the end, the metabolites
will be passed to another program which will identify those metabolites which are
identical or "similar" to known carcinogens. Those will be so marked in the
tree.

Presently, the major task is the aquisition of the metabolism knowledge
base, i.e. the writing of the transformation library to be utilized. Metabolism
experts at the National Institute of Health are gleaning this information from
both their own research and the metabolism literature. This information will be
encoded and the first testing of this new application for the SECS program will
begin in June 1977.

   E. Funding Status.

   Sandoz Unrestricted Grant to support Computer Synthesis $2500

   National Cancer Institute Contract   N01-CP-75815 "Computer-Aided Prediction
   of Metabolites for Carcinogenicity Studies" $56,328 for 18 months.

   Proposal RR-01059 submitted 1 March 1975 to Division of Research Resources,
   "Resource-Related Research: Biomolecular Synthesis", $227,816 for 3 years,
   approved 1 Oct 76, but still awaiting funding.

   Note: Were it not for the leased line and computer access granted to us by
SUMEX, the entire SECS project would not have been able to continue for the past
18 months.

   D. Current List of Project Publications

W.T. Wipke and P. Gund, "Simulation and Evaluation of Chemical Synthesis.
   Congestion: A Conformation Dependent Function of Steric Environment at a
   Reaction Center. Application with Torsional Terms to Stereoselectivity of
   Nucleophilic Additions to Ketones," J. Am. Chem. Soc., 98, 8107(1976).

W.T. Wipke, G. Smith and H. Braun, "SECS-Simulation and Evaluation of Chemical
   Syntheses: Strategy and Planning," ACS Symposium Proceedings, 1977.

W.T. Wipke, Computer Planning of Research in Organic Chemistry, Proceedings of
   the Third International Symposium on Computers in Chemical Education,
   Research, and Technology, Caracas, Venezuela, 1976.

S.A. Godleski, P.v.R Schleyer, E. Osawa, and W.T. Wipke, "The Systematic
    Prediction of the Most Stable Neutral Hydrocarbon Isomer," J. Am. Chem.
    Soc., 99, 0000(1977).

F. Choplin, R. Marc, G. Kaufmann, and W.T. Wipke, "Computer Design of Synthesis
    in Phosphorus Chemistry. Automatic Treatment of Stereochemistry," J. Am.
    Chem. Soc., 99, 0000(1977).

Manuals:

        SECS Users Manual, June 1976.
        SECS Users Guide, Aug 24, 1976.
        ALCHEM Tutorial, Sep 21, 1976.

II. Interactions with SUMEX-AIM Resource

    A. Examples of Collaborations and Medical use of Programs via SUMEX.
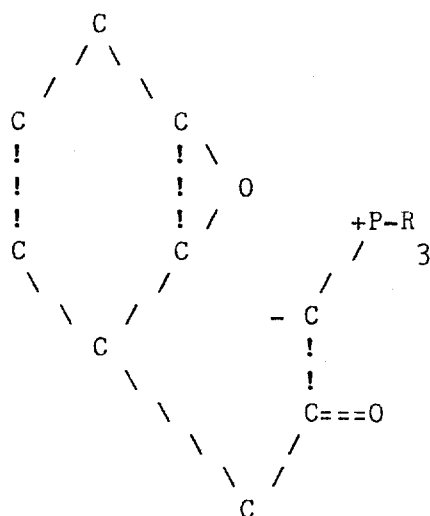
        SECS is available in the GUEST area of SUMEX and has been accessed
experimentally by many others as well. Professor R. V. Stevens (UCLA) explored
some syntheses of lycapodine while visiting Santa Cruz and as a result has
requested UCLA to obtain a graphics terminal so he and others at UCLA can access
SECS via SUMEX. Professor W. G. Dauben's group (Berkeley) has utilized the SECS
model builder on SUMEX is now extending the capabilities of that module of SECS.
Mr. Mel Spann of the National Library of Medicine toxicology program is
collaborating with us in developing a metabolism library for the metabolism of
catechol amines. Also collaborating with us on metabolism are Drs. Ted Gram from
Guarino's lab, Harry Gelboin, Dhiren Thakken and Harukiko Hagi from Jerina's lab,
Lance Pohl from Gillette's lab, Sidney Nelson from Mitchell's lab, Lionel Poirier
from Weisburger's lab, and Ken Chu and Sidney Siegel all of whom are from the
National Cancer Institute. Dr. Steve Heller of the EPA and Dr. G.A. Milne of
the National Heart and Lung Institute have expressed interest in putting SECS on
the Cyphernetics network as a part of the NIH chemical information system.
Restrictions on the allowed core image on that system have so far held up the
negotiations.

        For the past two years SECS has been available over TELENET from First Data
Corporation and has been accessed by industry: Squibb; Merck, Sharp and Dohme;
Pfize; Searle; Lederle Labs; FMC; and recently 3M Corporation and Stauffer. Dr.
Beryl Dominy of Fizer recently presented a paper before the Pharmaceutical
Manufacturer's Association entitled "SECS and the Information Scientist" in which
he describes his experiences with SECS, including an example where a synthetic
chemist was having difficulty with a particular synthesis, he then went to SECS
for possible solutions. SECS suggested another route as being better and indeed
that is what he found when he tried it later in the lab.

        The availability of SECS on SUMEX-AIM has also served health-related
research at the University of California, Santa Cruz. Model building using the
SECS model builder is being performed for Professor Edward Dratz (UCSC) to
generate conformations of fatty acids isolated from visual membranes ("Structure

and Function of Visual photoreceptors," EI00175), and for Professor Howard Wang
(UCSC) to study how conformations of steroids may affect the local anesthetic -
membrane interaction ("Role of Membrane Proteins in Local Anesthetic Action,"
GM22242).

We have assisted Professor J. E. McMurry in his synthetic work towards
Aphidicholine and Digitoxigenin by using the model builder for predicting
possible reaction pathways. An example is given below, where the conformation of
the epoxy-ylide was calculated along with the strain energies of the two possible
closure products.

```
                C
               / \
              /   \
           C        C
           !        ! \
           !        !  O
           !        ! /      +P-R
           C        C      /   3
            \      /      /
             \    /    - C
              C          !
               \         !
                \      C===O
                 \     /
                  \   /
                   C
```

Utilizing the SECS model builder, we have shown that attack on the epoxide
to form the fused system should be much more favorable then attack to form the
bicyclo compound. Similar studies have been undertaken to predict the
stereochemistry resulting from the acid catalyzed cyclization of McMurry's
Digitoxigenin precursor (HL-18118 "Total Synthesis of Cardiac Aglycones."):
application of SECS using a special library of cationic sigmatropic rearrangement
transforms generated the possible products which facilitated identification of
some of the side products in the early cyclization experiments. We have also
collaborated in the biogenesis work with Professor Phil Crews (UCSC) in marine
natural product biogenesis. Dr. Wipke has also used several SUMEX programs such
as CONGEN in his course on Computers and Information Processing in Chemistry.

B. Examples of Sharing, Contacts and Cross-fertilization with other
   SUMEX-AIM projects.

In collaboration with Dr. Ray Carhart and Dr. Dennis Smith of the
DENDRAL/CONGEN Project, a Computers in Chemistry Workshop was held at U.C. Santa
Cruz on the weekend prior to the Fall 1976 American Chemical Society National
Meeting held in San Francisco. The workshop attracted participants representing
all parts of the chemical community, academia, industry and government. Morning
lecture/discussion sessions introduced the SECS and CONGEN programs running on

SUMEX and the afternoon and evening sessions allowed "hands-on" experience for the participants. The response of the workshop participants was a very positive one with many participants showing so much interest that future collaboration and/or use of the powerful non-numerical computing tools available on SUMEX was discussed.

The SECS project has held joint research group meetings at Stanford with the DENDRAL and AI groups to discuss common problems and research goals. This has been very rewarding since the groups are complementary in orientation. These joint meetings also let the members meet in person after having met on-line on the network. Last year's AIM Conference at Rutgers was also a valuable experience, which allowed us to meet people interested in similar problems in different disciplines. It was particularly useful to have the opportunity to talk with experts designing new languages for knowledge representation and to hear them compare their systems.

C. Critique of Resource Services.

We find the SUMEX-AIM network very well human engineered. The ability to leave messages on the network, and to LINK to other users on-line for advice has been extremely useful to us since we have only the network to keep us educated about what is changing on the system, etc. The fact that we have been able to get productive research accomplished remote from Stanford speaks well for the SUMEX-AIM concept.

The SECS project finds the SUMEX-AIM staff and community extremely helpful, and anxious to extend themselves to meet our needs. SUMEX provided a leased line and modems to us and provided TYMNET access as well. Were it not for SUMEX, this research effort would have perished since there is no adequate computer facility on the UCSC campus or even in the UC system.

The only problems we have experienced are 1) until recently we were short of disk space, and 2) response time during the day can get pretty bad at times, particularly when using interactive graphics, so consequently most interactive graphics work is done at off hours. Basically we have found that SUMEX-AIM provides a productive and scientifically stimulating environment and we are thankful that we are able to access the resource and participate in its activities.

III. Follow-on SUMEX Grant Period (8/78-7/83)

A. Long-range User Project Goals and Plans:

Over this period of time the SECS project will continue research aimed at synthesis design and planning. Areas of interest include the formation of high level plans to guide the detailed chemical analysis, the capability for depth-first analysis, the evaluation of proposed synthetic pathways by forward simulation, and bidirectional search from target to key intermediate. At some time during this period the SECS program should be reimplemented in MAINSAIL to allow renovation of the SECS control structure and allow more machine