(Engelmore) has been debugging one of the basic analytical programs, the Rossmann rotation function, and is seeking further help from one of the protein crystallographers (Freer) at UCSD, who is logged in to SUMEX and has linked his terminal to Engelmore's. The transcript begins shortly after both parties have been notified that their terminals are linked. (Freer's comments are in lower case, Engelmore's in upper case preceded by a semi-colon.)

LINK FROM FREER, TTY 77

@;HI, STEVE. WHAT'S UP?

have you tried a general rotation (about a non-crystallographic axis)?

@;NO, I WANT TO DO THAT NOW, AND NEED SOME ADVICE ABOUT SETTING
@;IT UP. I WAS THINKING OF TAKING THAT LITTLE DI-PEPTIDE,
@;ROTATING IT ABOUT SOME AXIS BY SOME ANGLE, ADDING THE NEW SET
@;OF COORDINATES TO THE OLD SET, ESSENTIALLY MAKING A NEW
@;MOLECULE WITH TWICE AS MANY ATOMS, AND COMPUTING A NEW FC FROM
@;THERE.

all right, that is a good idea. it will check the capability of the program to locate non-crystallographic symmetry axes. remember to increase the size of the unit cell to accommodate the additional atoms.

@;LET ME SHOW YOU SOME OF THE COORDINATES AND THEN ASK A
@;QUESTION.
@TY THREON.ATM

; <ENGELMORE>THREON.ATM;3   THU 13-MAR-75 11:53AM   PAGE 1

TITLE L-THREONYL-L-PHENYLALANINE-P-NITROBENZYL ESTER
       HYDROBROMIDE

CELL	THRE	8.93	45.75	5.05	90.	90.	90.
GRID		10	10	10			
ATOMG	BR	4.563	0.363	-2.124	3.8	BR	
ATOMG	O	-1.867	0.274	-2.971	6.6	O1	
ATOMG	O	1.609	0.342	2.928	4.2	O2	
ATOMG	O	0.867	1.138	6.459	5.2	O3	

@;THERE ARE 30 ATOMS IN ALL. FIRST, SHOULD I DOUBLE THE CELL
@;SIZE?

i think you will be ok if you must double the size of the cell.

@;DO YOU HAVE ANY SUGGESTIONS FOR THE DIRECTION OF THE ROTATION
@;AXIS ABOUT WHICH I SHOULD ROTATE 180 DEGREES?

sorry for the delay, i am doing a couple of other things at the same time. you are more up on general rotations than i. you could make it parallel to a crystallographic axis if you wish (for ease of computation). just don't let it coincide with a

crystallographic axis.

@;OK, I THINK I WILL FIRST TRANSLATE THE MOLECULE BY SOME
@;"RANDOM" AMOUNT IN THE X-Z PLANE, THEN ROTATE IT 180 DEGREES.
@;OR VICE VERSA.

sounds ok.
are you going to perform the general rotation and increase the
cell size and then run a self-rotation rotation function run
while i am on line?

@;YES, LET'S DO IT. FIRST I'LL FIX UP THE CELL COORDINATES IN
@;THAT FILE I JUST LISTED. YOU CAN JUST WATCH FROM HERE ON ,
@;BUT FEEL FREE TO INTERRUPT AT ANY TIME.

(E. SETS UP AND RUNS A UTILITY PROGRAM TO GENERATE NEW ATOM
COORDINATES. THE OUTPUT, WHICH IS PRINTED ON BOTH E'S AND F'S
TERMINALS, IS OMITTED HERE FOR THE SAKE OF BREVITY.)

@;OK, NOW I HAVE TWO FILES OF ATOM COORDINATES TO MERGE. I DON'T
@;HAVE ANY AUTOMATIC WAY TO SEE IF THEY'RE OVERLAPPING, SO I
@;GUESS I'LL HAVE TO INSPECT THEM VISUALLY.

i wouldn't worry overly much about it. you have increased the
volume of the unit cell by a factor of 8 and the contents by 2.
a simple way to check is to sort on xyz and examine delta xyz of
adjacent pairs and list the pair with the smallest delta. don't
worry about it now, let's just go on and run the rot. program.

@;OK, WILL DO.

(E. SETS UP INPUT FOR THE PROGRAM WHICH COMPUTES THEORETICAL
STRUCTURE FACTORS.)

@;STEVE, SHOULD I GENERATE THE FC IN P1 INSTEAD OF P2(1)2(1)2(1)?

no, you have things set up for p2(1)2(1)2(1) now, so just stay in
that space group if you wish. the overlap problem would be
completely avoided if you switched to p1, but i don't think it
is necessary to do that now.

@IDATA2

(E. RUNS THE IDATA2 PROGRAM TO GENERATE STRUCTURE FACTORS. WHILE
THE PROGRAM IS RUNNING AND PRINTING INTERMEDIATE RESULTS ON BOTH
TERMINALS, AN ERROR IS DETECTED BY THE PROGRAM:)

TOO MANY REFLECTIONS

END OF EXECUTION
CPU TIME: 56.22 ELAPSED TIME: 3:2.91
EXIT.

@;TOO MANY REFLECTIONS????

remember, you doubled each cell edge and hence increased the volume of the cell by 8. you will therefore get 8 times as many reflections for a given minimum bragg spacing.

@;SO I GUESS I SHOULD ALSO REDUCE THE RESOLUTION FROM 2.0 TO 4.0,
@;RIGHT?

yes, that will reduce the number of reflections by a factor of 8.

@;OK, HERE WE GO AGAIN.

V. Proposed Extensions to Scope of Project vis-a-vis SUMEX Use.

Some of the components of the structure elucidation system under construction here are standard crystallographic analysis programs, used routinely by the UCSD protein crystallographers. These programs were originally written for the CDC 3600 at UCSD. Although that machine will continue to be their primary computing facility, the UCSD crystallographers have expressed an interest in performing some of their data reduction and other computational work on SUMEX, because of its larger core size and shorter turn-around times. The ARPANET connection now makes this activity more attractive, since relatively large files of data must be transferred between UCSD and SUMEX. This extension of the use of SUMEX for "standard" crystallographic computations would be beneficial to the Stanford side of the collaboration as well, by coupling the two groups more tightly through mutual development and use of software. As the UCSD crystallographers perform more of their computations at SUMEX, the locus of interest in crystallographic computing will shift towards remote computing via networks, providing a further demonstration of the utility of the SUMEX concept.

IV.A.2   NATIONAL USERS

IV.A.2.a   DIALOG PROJECT

Principal Investigators: Dr. H. Pople and J. Myers, M.D.
                           University of Pittsburgh

(Grant HEW MB-00144-01, 3 years, $167,168 this year)

I. ABSTRACT

DIALOG is a computer-based system for general medical
consultation that incorporates a hypothesis-formation model of
diagnostic logic, and an extensive medical data base now encompassing
approximately half of the major diseases of internal medicine. The
system has been designed to deal with complex clinical problems,
confounded by data produced by two or more distinct
clinicopathological entities. In dealing with such cases, the DIALOG
heuristic focuses successively on various aspects of a problem,
disregarding, as it does so, findings that are irrelevant in each
context. the system thereby exhibits diagnostic behavior comparable
to the problem-oriented approach of the skilled clinician.

II. SUMMARY OF PROJECT ACTIVITIES

When the DIALOG project was moved to SUMEX-AIM last fall, our
first priority was to convert the existing DIALOG program and data
base from LISP 1.6 to INTERLISP. This was accomplished without great
difficulty, and a working version of the system was operational on
SUMEX-AIM by late November.

At about the same time, an interactive data-entry system was
devised so as to enable expansion and refinement of the data base. As
an expedient, because the planned volume of data could not be
accommodated within the user address space of INTERLISP, this system
was implemented as an assemblage of five interacting INTERLISP forks.
While serving its intended purpose, which was to permit a twofold
expansion of the data base, this system proved unwieldy and costly in
its utilization of system resources.

Design of a successor system that could provide efficient access
to the newly expanded data base was begun in january of this year.
While continuing with INTERLISP as the host processor, we decided to
structure the entire vocabulary and network of associations comprising
the DIALOG data base in a set of disk files, with pages to be mapped
into a resident core buffer on a demand basis. This design required
that data management, core management, data entry and editing routines
be written in basic assembly language; coding and checkout of these
program modules is ongoing at the present time.

Plans for field test of the DIALOG system, outlined in our

original SUMEX-AIM proposal, have been held in abeyance pending completion of this system design and implementation effort. It is now projected that the testing and evaluation activities will commence in early fall of this year.

III. INTERACTION WITH SUMEX-AIM

We have found SUMEX-AIM to be a very habitable environment. The computer resources, both hardware and software, amply serve the present needs of the project and despite lack of local TENEX expertise, we have been able to accomplish a number of sophisticated systems programming tasks, thanks to superb documentation and a most helpful SUMEX-AIM staff.

Our only concern is that the SUMEX-AIM resource may have become overloaded, as the afternoon response time has degraded badly during recent weeks. We would hope that the physicians who are invited to aid in the evaluation of DIALOG this fall might be encouraged to exercise the system at their convenience, rather than at the convenience of the system scheduler.

IV.A.2.b   DISTRIBUTED DATA BASES FOR CHRONIC DISEASES

Distributed Data Base System for Chronic Diseases

R. A. Nordyke, M.D.; F. Kuo, Ph.D.; C. Kulikowski, Ph.D.

(Grant application in preparation)

Beginning in March 1975, we have begun to develop prototype consultation programs in chronic diseases that are to be linked to a set of distributed data bases. The collaborating institutions are Pacific Health Research Institute (R. A. Nordyke), University of Hawaii ALOHA System (F. Kuo) and Rutgers University (C. Kulikowski).

The initial phase consists of establishing a data base on thyroid diseases using already accumulated data. This will serve as a prototype for those in other chronic diseases such as hypertension and diabetes.

In developing computer programs for consultation, decision support and data acquisition and analysis, we recognize that different levels of complexity are appropriate depending on the particular patient's characteristics, the health care provider's role, and the purposes and environment of the encounter. To achieve a multiplicity of goals and provide a variety of clinical functions at different levels of sophistication it is most appropriate and effective if we distribute our work between local processors and shared national resources. In this manner we can best use of the specialized capabilities of each type of computer system and data base.

To satisfy local needs and preferences, we intend to design data acquisition and medical decision protocols to be run on small local minicomputers. These local minicomputers are to be linked via computer-communication networks (ALOHA and ARPANET) to the SUMEX computer at Stanford. Input/output control programs will be designed to facilitate modular and standardized transmission of important data that enter in the chronic disease data base. The ALOHANET can make the consultation programs available via radio-links and satellite to other Hawaiian islands and remote areas of the Pacific basin, where nurse/paramedic protocols for the management of chronic diseases could have a significant effect on the quality of health care delivery.

One of the principal objectives of our proposed research program is the development of sophisticated computer programs based upon methods of artificial intelligence to aid in the diagnosis and treatment of specific chronic diseases. This aspect of our work will build on already established collaborations between the clinical group at the Pacific Health Research Institute/Straub Clinic and the computer science group on artificial intelligence in medicine (AIM) at Rutgers University. Work at Rutgers (in collaboration with the Mt. Sinai School of Medicine) has concentrated on another chronic disease -- glaucoma. Thus, the proposed work in thyroid diseases represents a broadening of scope within a class of similar problems.

The systems currently under development possess unique logical and inferential capabilities for the explanation and justification of their decisions and strategies, based on their use of rational models of normal body function, disease processes, and patterns of health needs and health care. Because our models allow explanation of the process under study at many levels of complexity, they can generate information for many levels of users. For example, within a consultation program, diagnostic reasoning can be stressed for inexperienced practitioners, while prognosis and recommendations for therapy can be emphasized for those bearing greater responsibility for patient care.

The consultation programs within SUMEX will provide: multiple modules of disease, varying levels of resolution, different modes of interpretation (causal, logical, taxonomic, associative, probabilistic, etc.) facilities for explanation, instruction and querying of a data base of existing cases. A control program at the central resource (SUMEX) will receive requests from the local clinical control program and decide on the appropriate level and scope of the response. The local minicomputer will limit the type of information transfered to the resource machine to maintain security and confidentiality of the medical information. Local mass data will be on disc and on a large time-sharing computer, the BCC 500 at he University of Hawaii.

The data base for a specific chronic disease is best established at a clinical node where the appropriate knowledge and experience is available. Access to the data base by other investigators can be attained either through SUMEX (for AIM researchers) or directly through one of the computer networks..justify

IV.A.2.c   HIGHER MENTAL FUNCTIONS MODELING

HIGHER MENTAL FUNCTIONS MODELING (HMF)
Project Summary - April, 1975

Kenneth M. Colby, M.D.
Professor of Psychiatry, UCLA

(Grants NIH MH-06645-13, 3 years, $170,000 this year
and NIH MH-27132-01, 2 years, $130,000 this year)

Only since SUMEX has come onto the ARPA network have I been able
to work on the machine. Hence, there is not much to report.

1) GOALS

We plan to construct, test and validate an improved version
of a computer simulation of paranoid processes. This model has
clinical implications for the understanding, treatment and
prevention of paranoid disorders.

2) ACTIVITIES

The interactive model is now running on SUMEX, and we are
collecting data from the interviews to improve the model's
language recognition capabilities.

Now that SUMEX can be reached from UCLA, we expect to do a
great deal of work on the facility, particularly in debugging the
improved version of the model. Also, we intend to conduct some
experimental tests of the model in which experienced clinicians
rate its behavior along specific dimensions.

No publications as yet.

3) SUMEX INTERACTION

I have a lot of trouble getting on the machine. The
messages are "Timed Out", "Host Dead" and/or "Connection Open",
but nothing else happens. Also, the machine seems to go down a
lot during the late morning and early afternoon.

[NOTE: "These difficulties evidently reflect early problems with
reliability of ARPANET connection software and hardware." J.L.]

IV.A.2.d   MEDICAL INFORMATION SYSTEMS LABORATORY

Principal Investigators:

                    Dr. B. McCormick and M. Goldberg, M.D.
                    (Univ. of Illinois at Chicago Circle)

        (Grant HEW MB-00114-01, 2 years, $380,619 this year)

(1) OBJECTIVES

        The Medical Information Systems Laboratory (MISL) pursues three
major activities: clinical research and decision support; construction
and modeling of a data base in ophthalmology; and network-compatible
data base design. The priorities among these are such that the latter
two activities are ancillary to the exploration of artificial
intelligence techniques in clinical decision making. The resource
utilizes the computer facilities of the University of Illinois and the
SUMEX-AIM network, and provides the administrative structure for
assembling the expertise of the Department of Information Engineering
and the Department of Ophthalmology of the Illinois Eye and Ear
Infirmary.

        Clinical decision support activities are currently proceeding in
the following specific areas:

  a. Disease process nets: the investigation, development, and testing
     of knowledge-based models of disease processes.

  b. Data understanding systems in clinical medicine: development of a
     computer laboratory health care resource for the capture and
     validation of data at its source. The idea is to augment data
     capture with a clinical decision support model of what to seek,
     what to expect, and how to interpret data sets.

  c. Synthesis of optimal protocols: development of a formal system,
     called a "variable-valued logic system," with a few simple
     operations which -- when supplied with the observed data sets and
     the relations between the observed variables -- produces a
     description which is minimal. The system will be used to deduce
     the simplest rule for recognizing one class in the limited context
     of another class (or classes) of objects, and so prescribe the
     protocol for a differential diagnosis.

  d. Physician-guided decision support: an effort to provide the
     clinician and clinical researcher with tools for validating his
     data base, comparing his decision-making protocols with those of
     his colleagues, and initiating epidemiological studies.

  e. Support for epidemiological and longitudinal studies: involves
     interactive graphics access to classical biostatistical support
     programs; analysis procedures as in automatic control theory; and

risk and health hazard appraisals on the basis of current
epidemiological studies.

f. Investigations into formal aspects of modeling and knowledge
representation.

g. Extensive study of existing systems for automated diagnosis.

h. Creation of a natural language question-answering system for
extracting information from disease model data bases.

i. Sponsorship of series of seminars in relational data base design
and applied epistemology.

(2) PROGRESS

In MISL's first year most attention has been given to assembling
necessary equipment, personnel, and ideas. SUMEX-AIM was used for the
following purposes:

a. to facilitate a substantial review of literature on disease
modeling, formal modeling, and knowledge representation.
Information obtained over SUMEX-AIM was a considerable help in
assembling the relevant material.

b. to communicate with other groups engaged in similar activities,
especially the Computers in Biomedicine group at Rutgers.

c. for the construction of INTERLISP-based graphical language and
software for on-line graphics via a plasma panel.

d. to edit and generate intra-project memos and reports using such
SUMEX subsystems as SOS, TV-Edit, and PUB.

As MISL moves into its second year, its use of SUMEX-AIM is
expected to increase substantially. Program development will proceed
in the areas of disease modeling (of glaucoma, in conjunction with
Rutgers; and diabetic retinopathy), natural language query systems,
and systems for facilitating knowledge acquisition. SNOBOL and
INTERLISP should receive very heavy use.

(3) COMMENTS

Technically we have been well-served and well-pleased. The
system has proven to be easy to use and immensely powerful. The
services of Phil Jackson as consultant were very helpful in working
out some knotty system problems in the graphics project. The period
in which the system was quite unreliable was bothersome, as was the
period of long evening downtimes, but such initial troubles are to be
expected -- on the whole we have been very pleased with the
performance of the system.

The sharing of community computer and clinical resources (via the SUMEX-AIM network) is essential to MISL's vitality. From the beginning, collaboration with the Rutgers Computers in Biomedicine group has been a central consideration in MISL planning. We have been very significantly influenced by Casimir Kulikowski's and Sholom Weiss's ideas concerning the causal state modeling of disease. The Rutgers group has provided us with detailed information concerning their work, largely via the SUMEX-AIM network. Recently, Weiss has made his diagnostic program for glaucoma available over the network. This will be used in a clinical setting by physicians at the Illinois Eye and Ear Infirmary.

IV.A.2.e   RUTGERS COMPUTERS IN BIOMEDICINE

        Project: Rutgers Research Resource
                    Computers in Biomedicine

        Principal Investigator: Saul Amarel

        (Grant NIH RR-00643-04, 3 years, $285,240 this year)

I. PROJECT GOALS AND APPROACHES

        The fundamental objective of the Rutgers Resource is to develop
a computer based framework for significant research in the biomedical
sciences and for the application of research results to the solution
of important problems in health care. The focal concept is to
introduce advanced methods of computer science - particularly in
artificial intelligence - into specific areas of biomedical inquiry.
The computer is used as an integral part of the inquiry process, both
for the development and organization of knowledge in a domain and for
its utilization in problem solving and in processes of experimentation
and theory formation.

        The active Resource community consists at present of 26 members
and 12 collaborators. Members are mainly located at Rutgers.
Collaborators are located in several distant sites and they interact -
via SUMEX-AIM - with Resource members on a variety of projects,
ranging from system design/improvement to clinical data gathering and
system testing. At present, collaborators are located at the Mt. Sinai
School of Medicine, N.Y.; Washington University School of Medicine,
St. Louis, Mo.; Johns Hopkins Medical Center, Baltimore, Md.; Illinois
Eye and Ear Infirmary, Chicago, Ill.; College of Medicine and
Dentistry of New Jersey (CMDNJ); and the University of Hawaii -
Pacific Health Research Institute.

        Research in the Rutgers Resource is oriented to "discipline-
oriented" projects in medicine and psychology, and to "core" projects
in computer science, that are closely coupled with the "discipline-
oriented" studies. Work in the Resource is organized in four AREAS OF
STUDY; in each area there are several projects. The areas of study and
the senior investigators in each of them are:

    (1) Medical Modeling and Decision Making (C. Kulikowski, A. Safir).

    (2) Modeling Belief Systems (C. F. Schmidt).

    (3) Representations, Modeling and Hypothesis Formation in AI (S.
        Amarel).

    (4) Meta Description System (MDS) (C. V. Srinivasan).

In addition, the Rutgers Resource is sponsoring an Annual National AIM Workshop, whose main objective is to strengthen interactions between AIM activities, to disseminate research methodologies and results, and to stimulate collaborations and imaginative resource sharing within the framework of SUMEX-AIM. The Organizer of the first Annual Workshop (to be held at Rutgers on June 14 to 17) is C. Kulikowski; N. Sridharan is its Technical Director.

II. AREAS OF STUDY; SUMMARY OF GOALS AND ACTIVITIES

(1) Medical Modeling and Decision Making

   Present projects include:

   (i) Development and clinical testing of the Glaucoma Consultation program based on a Causal-Associational (CASNET) model - as a collaborative project of the Ophthalmological network which we have initiated last year (Mt. Sinai Medical School, Washington Univ., Johns Hopkins Univ. and Illinois Eye and Ear Infirmary).

   (ii) Investigation of models of disease description based on a general semantic network representation, with associated strategies of diagnosis, prognosis and therapy. These models subsume a variety of representations and sub-models useful in general consultation in ophthalmological diseases and selected chronic diseases. A particular emphasis is placed on the analysis of the true course of disease and interrelationships between various subprocesses.

   (iii) In collaboration with The Mt. Sinai Health Care Computer Laboratory we are developing models of refraction strabismus and neuro-ophthalmology.

   (iv) In collaboration with the Pacific Health Research Institute in Hawaii we are developing models of chronic diseases: thyroid disease, hypertension and diabetes.

   (v) In collaboration with CMDNJ we are developing a model in hematology.

   The following is a summary of accomplishments in this area:

   a. The Ophthalmological Network is functioning - consultation programs are now available through SUMEX-AIM to the four collaborating institutions.

   b. The Consultation System has been perfected by adding many details of diagnosis, pathophysiological states and new observations as the result of suggestions by the network participants.

c. A file system for storing cases and providing a chronological model-based interpretation has been created.

d. A set of programs to analyze the case histories is currently under development. When these are perfected they will be the initial step of a system for automatically incorporating experience into the consultation program.

This work is proceeding in close connection with an investigation in the AI area on models of grammatical inference and their relation to learning in causal structures of the type used in glaucoma.

The progress in the area of the Ophthalmological Network would be impossible without the facilities and support provided by SUMEX-AIM.

(2) Modeling Belief Systems

The overall goal of this project is to develop a computer-based psychological model of how persons reason about the causes of human action. The common-sense notion of social causation which is used to understand intentional actions has served as the focus of this effort. Within this paradigm, the observer explains the intentional actions of others by attributing to the others the plans and motives which could have generated the observed actions.

To date we have:

(a) defined the central concepts of such a notion of causality - that is, concepts such as person, act, plan, motive, belief, etc.

(b) identified and studied various strategies for reasoning about observed actions;

(c) investigated the implications of this type of reasoning for the organization of memory for such events;

(d) and investigated questions of how, in the child, these concepts of plans and motives might develop.

Our goals for the immediate future are:

(a) To continue our collaboration with the groups working in the AI and MDS areas in order to develop the model of social causation within the framework of MDS (the model is called BELIEVER);

(b) to continue the work on the study of strategies for reasoning about action as well as to develop empirical procedures that will aid us in the definition and study of such strategies;

(c) to continue our investigations on memory organization;

(d) to continue the work on how such causal reasoning develops in the child;

(e) to extend the theory of social causation to account for how persons evaluate, in a moral sense, the actions of others.

SUMEX-AIM is providing a common environment within which collaboration among the various investigators working on Belief Systems and those working on AI representation and MDS communicate and share programs. This has been particularly important during this year since the persons involved have been quite spread out geographically (Amherst, Mass., Cambridge, Mass., New Brunswick, N.J.).

(3) Representations, Modeling and Hypothesis Formation in Artificial Intelligence

A major part of our effort in this area is oriented to collaborations with investigators in other Resource projects - involving applications of AI ideas and programs and also identification and initial exploration of new significant AI problems.

Present projects include:

(i) Verifying consistency of a causal model of a physiological process (such as used in glaucoma) relative to an underlying mathematical model (N. Sridharan). Nlisp was the main tool used in this attempt. The current program provides a way of easily entering the mathematical model, performing checks for dimensional consistency of the variables entering into the equations, inferring dimensions of variables that were unspecified, using the graph depicting the relationships of variables and extracting the flow of determination in the variables. In order for the mathematical model to be consistent with a causal model, minimally one should ensure that the causal flow is compatible with the flow of determination.

(ii) Interactive acquisition of domain Knowledge in modeling Belief Systems (N. Sridharan). A prompting program was written to accept descriptions of act schemata written in a natural english-like syntax. The prompting system is general and will accept any structural description for which it will provide prompting. One has also the capability to specify transformations to be made on different fields of the structure.

(iii) Bringing up the FUZZY system of LeFaivre in INTERLISP and investigation of fuzzy problem solving approaches in medical decision making (R. LeFaivre). Implementation work is just starting here.

(iv) Application of grammatical inference schemes to automatic adjustment of medical causal models on the basis of clinical data (A. Walker). A survey of grammatical inference techniques was made, from the applications point of view, and the most promising technique was related to our causal modeling approach in the medical decision area. This led to theoretical results in the area of stochastic grammars - an area which promises to provide formal support for our work in medical modeling.

(v) Development of a grammatical inference system using a "developmental paradigm" (W. Fabens). This is a hypothesis formation system which attempts to change a given context free grammar so as to accommodate new sentences that cannot be derived from the given grammar. The system includes (a) a relaxation parser - which comes as close as it can to an interpretation of a given "deviant sentence", (b) a rule hypothesizer which uses such an interpretation to propose changes to the current grammar, (c) an intersection generator which tries to produce a sentence not legal before but legal after the rule modification process, and (d) a rule coalescer which summarizes with as little loss of gain in generality as possible the newly hypothesized grammar. We are currently concentrating on areas (c) and (d) and have developed programs for (a) and (b).

(vi) Development and study of systems for theory formation in programming tasks (S. Amarel). Experimental systems are being developed for cases where the program to be formed is specified in terms of (a) a desired output condition, (b) program traces for specific input-output pairs, and (c) a set of input-output pairs. The current approach to (c) is to make the generation of candidate programs (hypotheses) responsive to a detailed analysis of shortcomings of previously generated candidates. In this project, major emphasis is given to problems of representation and to the effects of shifts between representations.

This project and the previous two projects are focusing on different AI approaches to hypothesis (theory) formation - an area which is essential to the automatic acquisition and improvement of a Knowledge base from experimental data.

SUMEX-AIM is providing the LISP environment needed for developing some of the AI programs in this area, and (more importantly) it is providing an effective communication environment for collaboration between AI people and investigators in the Medicine and Psychology areas.

(4) Meta Description System (MDS)

MDS is a tool for building knowledge-based systems. It has two modes of operation. One is the domain acquisition mode. In this mode the system acquires the KNOWLEDGE in a domain, in a given description schema. The other is the domain execution mode. In this mode the system uses the described domain Knowledge automatically for problem solving in the domain. Considerable progress was made during the last year in building the system facilities for domain acquisition. The MDS system is being implemented in INTERLISP on the SUMEX-AIM computer.

We are now using the MDS framework to develop the system design for BELIEVER (the Belief Systems Model of Social causation). We have also described the GLAUCOMA consultation system in MDS.

III. AIM WORKSHOP

The theme of the first Annual Workshop is "Knowledge Based Systems in Medicine". The first day (June 14, 1975) will be devoted to a "General Session" which will provide an overview of current AIM activities and a broad forum for discussion. The following three days (June 15 to 17) will be devoted to discussions in depth of AIM designs, and to demonstrations of current systems.

The SUMEX-AIM system is essential for the Workshop. Most of the AIM programs will be running on SUMEX-AIM and accessed via TYMNET or ARPANET from Rutgers. The messages facilities of SUMEX-AIM have been most useful for planning, communicating and setting up the information pool for the AIM Workshop.

IV. EXPERIENCE WITH THE SUMEX-AIM RESOURCE

In the last year, we have used the SUMEX-AIM resource for program development and testing and for communication between investigators.

INTERLISP, SITBOL, FORTRAN, Editing and Message Handling systems were extensively used. The message and linking facilities were used to provide a common environment for collaboration between several investigators in a project (some of whom were quite spread geographically). For example, in the BELIEVER project we have established a common MSG file that is shared for READ/APPEND access by workers in the project.

SUMEX-AIM has provided the linkages (communications, terminals) for the establishment of the network of collaborators in ophthalmology. The initiation of the net was delayed because of problems with getting prompt delivery of equipment from manufacturers.

The SUMEX-AIM Staff has been most helpful and cooperative. The system has been very useful to us in establishing collaborative research activities, in sharing information, and in providing a forum for linking and talking.

As remote users, we are especially sensitive to the communication facilities available to SUMEX-AIM. The in-wats line performed well. TYMNET has improved since its early days; however, its performance for users in the New York region (the Mt. Sinai group) remains relatively poor. The ARPANET connection is now providing a good communication medium. With the linking of both SUMEX-AIM and RUTGERS-10 on the ARPANET, we are now in a better position to work effectively on both systems (file transfers,communications). Plans are now underway to make the RUTGERS-10 more compatible with SUMEX-AIM, so that it can provide a more reliable and convenient network environment for our investigators and their collaborators. S. Levy is coordinating these plans.

A source of annoyance with SUMEX-AIM has been the frequency and duration of downtime. Improvements in this area are of special significance for our research collaborations - especially in interactions with our medical collaborators. From the point of view of program developers, the system appears now frequently overloaded. When the load average shoots to 10, response time is extremely poor and useful work becomes impossible. This is another area, where the availability of a compatible RUTGERS-10 system may help.

In conclusion, the SUMEX-AIM facility is now very much an integral part of our research environment. Several important components of our project are completely dependent on it. Taking into consideration that this has been a year in which the system was being brought up, and many changes have been taking place, the SUMEX-AIM resource has provided a very fine support for the Rutgers project.

IV.B   INFORMAL PROJECTS

The following is a summary of the various "pilot" projects which
have been admitted to SUMEX on a temporary basis pending development
of a formal proposal. Many of these projects reflect initial efforts
at formalizing analyses of experimental situations in preparation for
the development of DENDRAL-like heuristic inference generation and
modeling.

IV.B.1   STANFORD PILOT PROJECTS

IV.B.1.a   ARTIFICIAL INTELLIGENCE APPLICATIONS IN GENETICS

Investigator: Prof. Cavalli-Sforza (Genetics)

(Grant NIH GM-20467-02, 3 years, $49,092 this year)

The following are reports from members of Dr. Cavalli-Sforza's
group who have used SUMEX this past year.

WAGENER

I have been using SUMEX primarily for REDUCE [A LISP system for
manipulation of symbolic algebraic expressions developed by A. Hearn
at Utah and transmitted to us via the ARPANET]. My research in
theoretical population genetics involves formulation of sometimes
quite complex mathematical models to describe various processes going
on within and between human populations which may affect the genetic
bases (genotypes) or expressions of these genotypes (phenotypes) in a
population.

Part of my PhD thesis includes the analysis of the evolution of
certain phenotypes that are influenced by interactions taking place
between both genetic and environmental factors. The object of this
research was to take a simple kind of interaction and show how
tendencies for mating assortatively affect the evolution of these
traits. When the mates are chosen nonrandomly, based on phenotypes,
the family environments expressed in the population are affected.
Then, when the distribution of expressed environments is changed, the
distribution of phenotypes affected by these environments may also
change. In the reduced form the recurrence equations of these models
are quite complex, sometimes involving as many as 100 terms. I found
little agreement between repeated attempts to verify the algebra by
hand. By modifying the equations slightly I was able to verify
conclusively these equations on REDUCE and was then able to program
the model.

AMMERMAN

We have been involved in the anthropological excavation of
several sites in Calabria, Italy. The data will give information
about the settlement patterns of Neolithic man as part of our analysis
of biological-anthropological and cultural adaptations. Correlation
of sites involves the use of several types of data: coordinates,
material found, period of material, elevation, land form, source of
nearest water, soil, geology classification and present land use. The
second phase of data collection is in progress as a preliminary to
more comprehensive modeling efforts.

THOMPSON

I have attempted to make a comprehensive simulation to generate
data on the evolution of chromosomes in a finite population. The
simulations are stochastic and involve mutation, many types of
selection, epistasis, and recombination. The affects of finiteness on
the evolution of certain traits (drift) may be studied. Also the
distribution of ages of mutants and selection on new mutants is
generated. The model is written in FORTRAN.

Investigator: Prof. J. Lederberg

Other Genetics Dept. Projects: Molecular Genetics
   & DNA Segments

My own laboratory group has been using SUMEX to help start the
development of new A-I programs for hypothesis-formation, automated
explanation and assistance in the induction and planning of new
experimental procedures in our laboratory work in molecular genetics.
We have therefore been using SUMEX wherever it was feasible for
supporting our day-to-day laboratory work AND where this also
contributed to establishing the requisite knowledge base.

The modelling of DNA subjected to segmentation by restriction
endonucleases at randomly occurring specific sites of bacterial DNA
has already been invaluable in the understanding of our experimental
findings. Cf:

Harris-Warwick, R., Ehrlich, S., Elkana, Y., & Lederberg, J.,
"Fraction and Purification of Bacterial Genes by Segmentation of DNA
with EcoR1 Endonuclease and Agarose-gel-electrophoresis", Proc. Nat.
Acad. Sci., U.S., In Press [June '75].

IV.B.1.b   INFORMATION PROCESSING PSYCHOLOGY PROJECT

Principal Investigators: Prof. E. Feigenbaum (Computer Science)
                            and Prof. H. Cohen (U. C. San Diego)

        (Grant application in preparation)

I.   Abstract of Project Goals and Activities

        The general goal of this research is the development of
information processing models of human problem solving, learning, and
memory, using techniques of computer simulation. In the spirit of much
previous work of this type by the artificial intelligence research
community (some of which has been sponsored by NIH and NIMH), this
work is to be thought of as application of AI concepts and techniques
to theory construction in Psychology. The effort currently most
active within the IPP Project is a pilot study of perceptual, memory,
and performance processes involved in certain types of construction of
visual forms (involving composition and production of freehand line
drawings). It is being carried out in collaboration with Prof. H.
Cohen, Dept. of Visual Arts, Univ. of Calif., San Diego.

II. Summary of Project Accomplishments

        The primary activity of the IPP project during the year has been
the pilot study mentioned in I above. The "pilot" intention is to
prepare the work to be discussed below for formal grant proposal
submission to NIMH. If that project is eventually approved, it will
apply for separate status at SUMEX.

        The current study of art-making behavior is concerned with the
modeling of how significance (="meaning") is ascribed to symbols and
groups of symbols, and how symbols are manipulated for the generation
of significance.

        Previously, computer simulation models have been used to study a
subset of meaning which belongs in the domain of communication and
concerns itself with verifiable fact. Thus, problem-solving models are
expected to find answers which are demonstrably "right", inference
models to draw inferences which are verifiable, conversation models to
"understand" what the conversant intended to communicate. The reasons
for this bias towards "well-defined" problems are clear and
reasonable.

        The more general view adopted in this work involves the
propensity in human cognition to GENERATE meaning. Under study is the
view that "creative" modes of intellectual behavior may be thought of
as game-playing in the domain of meaning-generation.

        The project work breaks down into the following stages:

1. The delineation of a group of drawing protocols, the exercise of which will trigger the meaning-generating propensity of the viewer; that is, which will present the viewer with an invitation to play with the significance of the drawing. In order to demonstrate the existence of this function in the viewer, and its importance in the whole "meaning" transaction, it is required that the model exhibit no intentionality with respect to meaning.

2. Once the separation of (viewer) generated meaning from (artist) intended meaning is established, it will be necessary to take into account that artists would normally exhibit intentionality - whether or not that intentionality is the source of "meaning" for the viewer - and that this intentionality must be expressed in relation both to the artwork and to the artist's view of the world. We will then need to consider in what way the artist's view of the world might function to provide determinants to his art-making behavior. This last question is particularly important, since art-making performance does not appear to be as goal-oriented as performance in chess, infectious disease diagnosis, or most other tasks studied by AI researchers.

The project is currently (predominantly) in Stage 1, though an effort has been made to develop a formalism which will support projected work in Stage 2 without monumental reworking. In order to accomplish this some preliminary work has been done on a simple associative memory model of a conventional semantic net form, and a program has been written which will permit the user to build up a large "sample" memory for testing purposes very rapidly.

The balance of the effort of the last few months has been directed to the clear definition of what a drawing protocol might be. Since this proceeded from the constraint of non- intentionality with respect to significance, the program has never at any stage adopted a lexical approach. In place of morphological units like lines and squares and circles the program performs with perceptual units like figure/ground differentiation and inside/outside differentiation, and with task units like pathfinding. Closed forms of considerable complexity are generated from the notion of closure, not from a lexicon of polygons. CLOSURE is thus an example of a protocol.

This protocol-oriented formalism has been generalized into a production system in which the left sides list states of protocol history and the right sides give appropriately weighted tables of admissible protocols. The production system thus embodies the model's explicit knowledge of image-making, and is in a form which should eventually permit extension to world knowledge.

The list of available protocols is now being developed vertically to include a number of REPETITION protocols, thus opening up the possibility of hierarchical performance; e.g. the application of the CLOSURE protocol repetitively to a recent protocol sub-history. (a ring of rings, etc.)

The replacement of the traditional lexicon-based morphology by a perception-informed drawing-protocol-based morphology promises to prove a powerful tool in the reexamination of a wide range of primitive image-making. A preliminary attempt is now being made to apply it - by hand - to a formal description of the Chalfont Valley (Calif.) group of Indian petroglyphs.

III. Comments; Assessments; Correlation with Expectations

The interaction with the SUMEX machine and the SUMEX organization has been eminently satisfactory; no change is expected or desired.

As is the case for most users, the need for more file storage is critical, but we have no special demands in this dimension.

We do not envision use of the system in excess of the usage previously discussed; plan no new departures that will involve significant computer use; and will remain in pilot project mode until submission and approval of our grant request later this year.

IV.B.1.c   AIM RESEARCH - UNIVERSITY OF ROCHESTER

Investigators: Drs. Feldman, Rovner, and Low
                Rochester University

        (Grant NSF DCR74-24203, 2 years, $149,956 total and
        Sloan Fdn. 74-12-5, 3 years, $120,000 this year)

        The Rochester group has a number of projects in connection with
the SUMEX-AIM facility. These range from the use by medical students
of existing systems to research on programming languages and
techniques for Artificial Intelligence. The group itself is quite new
and will be initiating new research programs over the next year as the
staff builds and our research plans develop sharper focus. This note
will describe existing programs and then briefly discuss anticipated
efforts. In addition to our research work on SUMEX, we have worked on
system development primarily in BCPL and SAIL.

A. Evaluation of Existing Systems

        Some second year medical students have been evaluating MYCIN
from the point of view of a physician under the guidance of Dr. Roy
Steigbigel.a specialist in infectious disease and Dr. Charles
Odoroff, Chairman of the Biostatistics program.

        The students have tried a number of cases based on their course
work and on hospital experience suggested by Dr. Steigbigel. There
has been some communication with Bruce Buchanan about the system
problems, so the beginnings of an interaction are there. An
assessment of the applicability of MYCIN to clinical problems is due
in early June. A copy will be forwarded when it is available.

B. Automatic Choice of Data Representation

        Abstract data structures such as sets, lists and relations are
being used more and more within programming languages for Artificial
Intelligence. They ease programming by providing good models with
which the programmer can express his problem. However there is no one
fixed representation of any such abstract data structure which is
optimal for all programs. Each set representation (such as linked
lists, binary trees, boolean arrays and so forth) has its own
properties. Depending on how a program uses an abstract structure,
one representation may be far superior to others in terms of storage
occupied by the data structure and/or the time needed for manipulating
it. For example, if the operations on a particular set are just
membership testing, insertion and removal of elements then a boolean
array may be the best representation. However if the size of the set
is normally small but the number of potential elements is large, this
representation is very expensive in storage requirements. If the set

is iterated through (FOREACH element in the set do some operation), the time needed for iterating through the set is proportional to the number of potential elements not the actual number of elements. Other representations have different storage and time requirements. The problem is to choose an appropriate representation for each set and list of SAIL program instead of having one fixed representation which is always used. A system for doing this was developed by Jim Low, now at the University of Rochester, as part of his dissertation research at Stanford. This systems has, in the last few months, been adapted to run under TENEX at SUMEX. This system uses information obtained by static analysis, of the SAIL program, dynamic statement execution counts, and information supplied by interrogation of the user.

We hope to continue to enhance this system by improving the decision making processes and extending the capabilities to allow change of representation.

I. More Efficient Implementation of LEAP

The associative aspect of LEAP of the programming language SAIL is a data base of 3-tuples. LEAP allows the programmer to query the data base in any of seven ways characterized by the those fields left unspecified for the search. In the following A, O, V represent specified fields of the 3-tuple.

(1) (A,O,V) (does a totally specified 3-tuple exist)

(2) (?,O,V) (find all 3-tuples with specific 2nd and 3rd components)

(3) (A,?,V)

(4) (A,O,?)

(5) (?,?,V)

(6) (A,?,?)

(7) (?,O,?)

The current implementation of LEAP uses hash-coding techniques to perform searches (1) and (4). An inverted file scheme (on the third component) is used to perform searches (2), (3) and (5). The data structure used in representing 3-tuples does not allow direct searches of types (6) and (7). The current implementation simply iterates through all possibilities of the first position for search (7) (all possibilities of the second position for search (6)) and then performs a search of type (4). Thus, with the current implementation, searches of types (6) and (7) are orders of magnitude slower that the other searches.