

IV. Description of Scientific Subprojects

The following subsections report on the AIM community of projects and "pilot" efforts including local and national users of the SUMEX-AIM facility at Stanford. In addition to these detailed progress reports, abstracts for each project and its individual users are submitted on a separate Scientific Subproject Form. However, we have included briefer summary abstracts of the fully-authorized projects in Appendix E on page 277.

Those groups from the National AIM community which use the SUMEX-AIM resource solely for communication (i.e., electronic mail to and from colleagues or access to bulletin boards and other information resources at SUMEX) are listed starting on Page 220, without detailed reports on their research.

The detailed collaborative project reports and comments are the result of a solicitation for contributions sent to each of the project Principal Investigators requesting the following information:

- I. SUMMARY OF RESEARCH PROGRAM
 - A. Project rationale
 - B. Medical relevance and collaboration
 - C. Highlights of research progress
 - Accomplishments this past year
 - Research in progress
 - D. List of relevant publications
 - E. Funding support
- II. INTERACTIONS WITH THE SUMEX-AIM RESOURCE
 - A. Medical collaborations and program dissemination via SUMEX
 - B. Sharing and interactions with other SUMEX-AIM projects
(via computing facilities, workshops, personal contacts, etc.)
 - C. Critique of resource management
(community facilitation, computer services, communications services, capacity, etc.)
- III. RESEARCH PLANS
 - A. Project goals and plans
 - Near-term
 - Long-range
 - B. Justification and requirements for continued SUMEX use
 - C. Needs and plans for other computing resources beyond SUMEX-AIM
 - D. Recommendations for future community and resource development

We believe that the reports of the individual projects speak for themselves as rationales for participation. In any case, the reports are recorded as submitted and are the responsibility of the indicated project leaders. The only exceptions are the respective lists of relevant publications which have been uniformly formatted for parallel reporting on the Scientific Subproject Form.

IV.A. Stanford Projects

The following group of projects is formally approved for access to the Stanford aliquot of the SUMEX-AIM resource. Their access is based on review by the Stanford Advisory Group and approval by Professor Shortliffe as Principal Investigator.

In addition to the progress reports presented here, abstracts for each project and its individual users are submitted on a separate Scientific Subproject Form.

IV.A.1. BBICU Project

BBICU Project: Blackboard Applications in the Intensive Care Unit

Adam Seiver, M.D.
Department of Surgery
Veterans Administration Hospital
Palo Alto, California

Lawrence Fagan, M.D., Ph.D.
Department of Medicine
Stanford University

Barbara Hayes-Roth, Ph.D.
Department of Computer Science
Stanford University

I. SUMMARY OF RESEARCH PROGRAM

A. *Project Rationale*

We are designing a data-interpretation and therapy-planning system for the intensive care unit (ICU). Fundamental research issues in temporal reasoning are associated with the ICU application area including assimilation of incoming data, representation of time-oriented intervals, and description of ongoing physiological processes [Fagan 84]. In addition, in ICUs of the 1990s, many more physiological measurements will need to be collected at frequent intervals, and increased access to the current medical record in coded format will be possible. Processing of incoming data will have to be opportunistic, selecting from a number of models that have different computational requirements and accuracy. We will use a blackboard architecture, known as BB1, that has evolved from earlier work on protein-structure elucidation and construction layout [Hayes-Roth 85]. BB1 is particularly well suited for the ICU project because it maintains separate blackboards for domain and control knowledge.

Although we have selected the blackboard structure as the organizing principle, many knowledge representation issues remain. First, we must represent the structure and function (anatomy and physiology) of the respiratory system. By characterizing the pathophysiology in terms of generic faults, we will create a more flexible means to diagnose problems in unusual situations -- in contrast to the phenomenological rules used in earlier systems.

Second, we must coalesce quantitative and qualitative models. The physiology of the respiratory and cardiac systems have been modeled in detail, but it is impractical to base the entire reasoning process on complex mathematical equations. Instead, we are developing methods to transform quantitative models into simpler formulations. We must make explicit the simplifying assumptions and associate them with their corresponding models, so that we can select an applicable model for situations of varying complexity.

Using the approach of our project for planning treatments for cancer patients [Langlotz 87], we will use strategic knowledge to create patient-specific specializations of standard treatment plans. We will use decision analytic methods to

evaluate and explain the various treatment options available at any point in time. The long-term goal of this project is to embed the decision-making components within the data management tasks of the ICU.

B. Medical Relevance and Collaboration

The problem of too much data being generated in the ICU is well recognized. Originally, monitors were designed to provide more objective assessments of the physiology of the patients in life-threatening situations. However, as more and more measurements became available, the ability of clinicians to assimilate the data began to drop. Expert systems can be designed to sort through the data, recognize untoward events in context and help with therapy selection. An early version of this was Fagan's VM system, which was based on extensions to the production rule framework. The current research has far broader goals, including real-time response using multiple methods for reasoning, reasoning from anatomy and physiology, and the use of integrated mathematical models. This has led to a three way collaboration between the VA hospital which is installing a data management system for a new Surgical ICU (Seiver), the Computer Science Department where blackboard research for real-time systems is in progress (Hayes-Roth), and the Medical Computer Science Group where investigations in qualitative-quantitative reasoning is taking place (Fagan).

C. Highlights of Research Progress

This project is in its early stages, but we are beginning to see progress in both research directions. The BB1 group has demonstrated a system that can reason by analogy from a set of symptoms corresponding to a physical "blockage" back to generic knowledge about structures involved in a flow process. This system has been developed in the BB1 architecture on the Explorer Lisp machine. This group has developed a prototype that can:

1. monitor a few types of electronically sensed data in real time,
2. dynamically focus attention on different types of data depending on the current situation,
3. incrementally classify asynchronously arriving data into temporal episodes of normal/abnormal physiological parameters,
4. dynamically compute conditional probabilities of alternative diagnoses as newly arriving data are classified, and
5. create alternative hypotheses when compiled clinical knowledge fails to explain the situation by reformulating the problem in terms of an underlying model of the structure and function of the body.

To date, these goals have been carried out with a small number of data streams and recognized diagnoses. A major research goal of this project is to understand how the knowledge representations and processing techniques will have to adapt as the problem is scaled up.

The quantitative/qualitative group has concentrated on building a variety of mathematical models based on the Dickenson model of oxygen transport through the circulatory system. These equations are at different levels of specification, and heuristics are used to select the most appropriate model at each point in time. Another active area of research is the relationship between the therapy planning module and the solution of mathematical models used for prediction.

D. Relevant Publications

1. Fagan, L., Kunz, J., Feigenbaum, E, and Osborn, J. Adapting a rule-based system for a monitoring task, in *Rule Based Expert Systems: The Mycin Experiments of the Stanford Heuristic Programming Project*, B. Buchanan and E. Shortliffe (eds.). Reading, MA: Addison-Wesley Publishing Co., 1984.
2. Hayes-Roth, B. A Blackboard architecture for control. (*Artificial Intelligence*) 26:251-321, 1985.
3. Langlotz, C., Fagan, L., Tu, S., Sikic, B., and Shortliffe, E. A therapy planning architecture that combines decision theory and artificial intelligence techniques. *Computers and Biomedical Research* 20:279-303, 1987.

II. INTERACTIONS WITH THE SUMEX-AIM RESOURCE

A. Medical Collaborations and Program Dissemination via SUMEX

As described above, this project is a three-way collaboration between the Departments of Computer Science and Medicine, and Department of surgery at the VA Hospital. As such we will need electronic mail and networking facilities. In addition, we imagine strong interactions with other projects around the world with similar research goals. We have already been contacted by research groups in Holland, Scotland, and Norway. In addition, similar research projects are underway at Yale, Berkeley, and Chicago. We expect that the networking facilities may allow us to share test cases, and possibly knowledge bases.

B. Sharing and Interaction with Other SUMEX-AIM Projects

The Yale project mentioned above is associated with Perry Miller's group. We also expect considerable interaction with the ONCOCIN and other parts of the Heuristic Programming Project at Stanford. The temporal issues involved with this project are relevant to Larry Widman. As the last AIM Workshop at Stanford in April, 1988, Larry demonstrated his programs on SSRG-based Explorers with the aid of SSRG staff member Rich Acuff.

C. Critique of Resource Management

The SUMEX staff have been quite helpful in the support of the various machines that have been used in this project so far. We anticipate that the project will migrate to Micro-Explorers and/or to Mac IIs. We believe that the current efforts of the SUMEX staff are quite appropriate for our research needs.

III. RESEARCH PLANS

Our basic research agenda is described above. The basic research issues underlying this project will extend for several years, leading to an implementation in the Veterans Administration Hospital in Palo Alto.

This research will continue to need help assistance with local area networking, file

service, and inter-network mail. We will need support for communications support within a project that is spread out over three geographical sites, and working with related but not identical hardware.

Since this is a new project, we are only depending on the DEC-20 for mail, file service, and communication facilities. The SUN-4 arrangement should be able to provide these services if local area network mail is fully implemented.

The SUMEX staff has been quite useful in providing support in other configurations of mainframe and workstations networked together. We anticipate that support for our unique collaborative arrangement will be equally superb.

IV.A.2. GUIDON/NEOMYCIN Project

GUIDON/NEOMYCIN Project

William J. Clancey, Ph.D.
Department Computer Science
Stanford University

Bruce G. Buchanan, Ph.D.
Computer Science Department
Stanford University

I. SUMMARY OF RESEARCH PROGRAM

A. *Project Rationale*

The GUIDON/NEOMYCIN Project is a research program devoted to the development of a knowledge-based tutoring system for application to medicine. The key issue for the GUIDON/NEOMYCIN project is to develop a program that can provide advice similar in quality to that given by human experts, modeling how they structure their knowledge as well as their problem-solving procedures. The consultation program using this knowledge is called NEOMYCIN. NEOMYCIN's knowledge base, designed for use in a teaching application, is the subject material used by a family of instructional programs referred to collectively as GUIDON2. The problem-solving procedures are developed by running test cases through NEOMYCIN and comparing them to expert behavior. Also, we use NEOMYCIN as a test bed for the explanation capabilities incorporated in our instructional programs.

The purpose of the current contracts is to construct a knowledge-based tutoring system that teaches diagnostic strategies explicitly. By strategy, we mean plans for establishing a set of possible diagnoses, focusing on and confirming individual diagnoses, gathering data, and processing new data. The tutorial program has capabilities to recognize these plans, as well as to articulate strategies in explanations about how to do diagnosis. The strategies represented in the program, modeling techniques, and explanation techniques are wholly separate from the knowledge base, so that they can be used with many medical (and non-medical) domains. That is, the target program will be able to be tested with other knowledge bases, using system-building tools that we provide.

B. *Medical Relevance and Collaboration*

There is a growing realization that medical knowledge, originally codified for the purpose of computer-based consultations, may be used in additional ways that are medically relevant. Using the knowledge to teach medical students is perhaps foremost among these, and GUIDON2 focuses on methods for augmenting clinical knowledge in order to facilitate its use in a tutorial setting. A particularly important aspect of this work is the insight that has been gained regarding the need to structure knowledge differently, and in more detail, when it is being used for different purposes (e.g., teaching as opposed to clinical decision making). It was this aspect of the GUIDON research that led to the development of NEOMYCIN, which is an evolving computational model of medical diagnostic reasoning that we hope will enable us to better understand and teach diagnosis to students. An important additional realization is that these structuring methods are beneficial for improving

the problem-solving performance of consultation programs, providing more detailed and abstract explanations to consultation users, and making knowledge bases easier to maintain.

As we move from technological development of explanation and student modeling capabilities, we are now collaborating closely with medical students and physicians to design an effective, useful tutoring program. In particular, medical students have served as research assistants, and a recent MSAI student is an experienced physician, John Sotos, from Johns Hopkins. The project has also collaborated with a community of researchers focusing on medical education, funded by the Josiah Macy, Jr. Foundation.

C. Highlights of Research Progress

C.1 Accomplishments This Past Year

C.1.1 The GUIDON-MANAGE Tutoring Program

This program teaches a student the language of diagnosis by having him or her enter all requests for patient information as an *abstraction*. Thus, the student issues "strategic commands" such as "test the hypothesis meningitis" or "ask a follow-up question about the headache," and the program (NEOMYCIN) carries out the tactics. By year end, this program was operational, with a complex interpreter for simulating NEOMYCIN to generate help, a feedback window to indicate what NEOMYCIN did when it carried out the commands, and many menus for making input to the program convenient. Research continues to focus on the assistance and feedback components of the program.

C.1.2 Explanation Research

Research in explanation is another major area. This year we adopted a new approach of printing the least possible information that would convey a line of reasoning, rather than generating text to describe everything the program did. The new explanation is based on the idea that the questioner seeking an explanation finds explanations more acceptable if it is necessary to fill in some gaps by his or her own effort. Our current approach is to print the domain relations (e.g., causes or subtype) that link foci to each other, leaving it to the questioner to infer the strategy behind each focus shift. This methodology will allow us to build up a more principled theory of explanation by determining what minimal information is acceptable and what causes an explanation to be inadequate.

C.1.3 The ODYSSEUS Modeling Program

Our third tutorial-related project involves continued development of a modeling program, ODYSSEUS. The purpose of ODYSSEUS is to discover domain knowledge discrepancies between an application domain knowledge base (e.g. the NEOMYCIN medical knowledge base) and a student or expert problem solver. IMAGE, an earlier modeling program developed in 1982, did not address this problem. The input to ODYSSEUS is the problem solver's patient data requests. When ODYSSEUS watches a student it functions as a student modeling program for GUIDON2 and when it watches an expert it functions as a knowledge acquisition program for HERACLES.

During the past year, a dissertation describing the program has been completed.

C.1.4 The HERACLES Expert System Shell

The final major effort involves generalizing our expert system tool, HERACLES, so that it can be made available to other research groups who wish to develop knowledge

bases which can be tutored by GUIDON2. Several copies of HERACLES were shipped on floppies in the past year to users of Xerox D-Machines.

C.1.5 Dissemination of Results

Besides publications, a number of tutorials and invited talks by Dr. Clancey presented this work around the world:

- Tutorial Speaker, Evaluation of Expert Systems, AAAI-87, July 1987.
- Tutorial Speaker, Second Advanced Course on AI, Norway, August 1987, six hours of lectures.
- Tutorial Speaker, Knowledge-Based Tutoring, IJCAI-87, Milan, August 1987.
- Invited Speaker, AI-87, Osaka, Japan, October 1987.
- Main Tutorial Speaker, "A Perspective on Knowledge Engineering," Inter Access, The Hague, Netherlands, February 1988, eight hours of lectures.
- "Intelligent Tutoring Systems," panel of the Cognitive Science Society, Seattle, July 1987.
- National Space Sciences Educational Foundation, Stanford, July 1987.
- Fujitsu, Tokyo Japan, CS Forum, November 1987.
- CSK/CRI, Tokyo Japan, CS Forum, November 1987.

C.2 Research in Progress

As of March, 1988, Dr. William Clancey has moved to the Institute for Research on Learning in Palo Alto, California, where he continues his research on teaching and learning. His use of the SUMEX resource is reduced to accessing archived files.

D. Publications Since January 1987

1. Clancey, W.J. Knowledge-Based Tutoring: The GUIDON Program, Cambridge: MIT Press.
2. Clancey, W.J. From Guidon to Neomycin and Heracles in twenty short lessons: ONR final report, 1979-1985. Current Issues in Expert Systems, 79-123, Academic Press, Inc., London.
3. Clancey, W.J. Intelligent tutoring systems: A tutorial survey. Current Issues in Expert Systems, 39-78, Academic Press, Inc., London.
4. Clancey, W.J. The knowledge engineer as student: Metacognitive bases for asking good questions. In Learning Issues in Intelligent Tutoring Systems, eds. H. Mandl and A. Lesgold, Springer-Verlag, in press. Also KSL 87-12.
5. Wilkins, D.C., Clancey, W.J., and Buchanan, B.G., Knowledge Base Refinement Using Abstract Control Knowledge. January, KSL-87-01.
6. Wilkins, D.C., Buchanan, B.G., and Clancey, W.J., *The Global Credit Assignment Problem and Apprenticeship Learning*. January, KSL-87-04.

E. Funding Support

Contract Title: "A Family of Intelligent Tutoring Programs for Medical Diagnosis"

Principal Investigator: Bruce G. Buchanan, Prof. Computer Science, Research
Associate Investigator: William J. Clancey, Research Assoc. Computer Science
Agency: Josiah Macy, Jr. Foundation
Term: March 1985 to March 1988
Total award: \$503,415 direct costs

Contract Title: "Computer-Based Tutors for Explaining and Managing the Process of Diagnostic Reasoning"

Principal Investigator: Bruce G. Buchanan, Prof. Computer Science, Research
Associate Investigator: William J. Clancey, Research Assoc. Computer Science
Agency: Office of Naval Research
ID number: N00014-85-K-0305
Term: March 1985 to November 1989
Total award: \$712,411 total

II. INTERACTIONS WITH THE SUMEX-AIM RESOURCE*A. Medical Collaborations and Program Dissemination via SUMEX*

We are frequently asked to demonstrate GUIDON-MANAGE, GUIDON-WATCH, and NEOMYCIN to Stanford visitors or at meetings in this country or abroad. Physicians have generally been enthusiastic about the potential of these programs and what they reveal about current approaches to computer-based medical decision making. We use network e-mail through SUMEX to communicate with other researchers worldwide.

B. Sharing and Interaction with Other SUMEX-AIM Projects

We interact periodically with Paul Feltovich at Southern Illinois Medical School. In addition, the central SUMEX development group acts as an important clearing house for solving problems and distributing new methods.

C. Critique of Resource Management

The SUMEX resources group has provided exemplary service. We have no complaints or suggestions whatsoever.

III. RESEARCH PLANS*A. Project Goals and Plans*

This research project has now moved from Stanford University to the new Institute for Research on Learning (IRL). We will no longer be an active member of the SUMEX Resource.

B. Requirements for Continued SUMEX Use

We have arranged to have archival access to our code and research notes (via dump tapes), which we prepared and stored at SUMEX from 1974 through 1987. We hope to move our personal archived files to our own disks in the next year, but will benefit from continuing access for project files over the next few years.

IV.A.3. MOLGEN Project

**MOLGEN - Applications of Artificial Intelligence to Molecular
Biology: Research in Theory Formation, Testing, and Modification**

**Prof. E. Feigenbaum
Department of Computer Science
Stanford University**

**Dr. P. Friedland
NASA-Ames Research Center
Moffett Field, CA**

**Prof. Charles Yanofsky
Department of Biology
Stanford University**

I. SUMMARY OF RESEARCH PROGRAM

A. Project Rationale

The MOLGEN project has focused on research into the applications of symbolic computation and inference to the field of molecular biology. This has taken the specific form of systems which provide assistance to the experimental scientist in various tasks, the most important of which have been the design of complex experiment plans and the analysis of nucleic acid sequences. Our current research concentrates on scientific discovery within the subdomain of regulatory genetics. We desire to explore the methodologies scientists use to modify, extend, and test theories of genetic regulation, and then emulate that process within a computational system.

Theory or model formation is a fundamental part of scientific research. Scientists both use and form such models dynamically. They are used to predict results (and therefore to suggest experiments to test the model) and also to explain experimental results. Models are extended and revised both as a result of logical conclusions from existing premises and as a result of new experimental evidence.

Theory formation is a difficult cognitive task, and one in which there is substantial scope for intelligent computational assistance. Our research is toward building a system which can form theories to explain experimental evidence, can interact with a scientist to help to suggest experiments to discriminate among competing hypotheses, and can then revise and extend the growing model based upon the results of the experiments.

The MOLGEN project has continuing computer science goals of exploring issues of knowledge representation, problem-solving, discovery, and planning within a real and complex domain. The project operates in a framework of collaboration between the Heuristic Programming Project (HPP) in the Computer Science Department and various domain experts in the departments of Biochemistry, Medicine, and Biology. It draws from the experience of several other projects in the HPP which deal with applications of artificial intelligence to medicine, organic chemistry, and engineering.

B. Medical Relevance and Collaboration

The field of molecular biology is nearing the point where the results of current research will have immediate and important application to the pharmaceutical and chemical industries. Already, clinical testing has begun with synthetic interferon and human growth hormone produced by recombinant DNA technology. Governmental reports estimate that there are more than two hundred new and established industrial firms already undertaking product development using these new genetic tools.

The programs being developed in the MOLGEN project have already proven useful and important to a considerable number of molecular biologists. Currently several dozen researchers in various laboratories at Stanford (Prof. Paul Berg's, Prof. Stanley Cohen's, Prof. Laurence Kedes', Prof. Douglas Brutlag's, Prof. Henry Kaplan's, and Prof. Douglas Wallace's) and over four hundred others throughout the country have used MOLGEN programs over the SUMEX-AIM facility. We have exported some of our programs to users outside the range of our computer network (University of Geneva [Switzerland], Imperial Cancer Research Fund [England], and European Molecular Biology Institute [Heidelberg] are examples). The pioneering work on SUMEX has led to the establishment of a separate NIH-supported facility, BIONET, to serve the academic molecular biology research community with MOLGEN-like software. BIONET is now serving many of the computational needs of over two thousand academic molecular biologists in the United States.

More generally, our work in qualitative simulation as applied to molecular biology is also relevant to building models of many other medical and biological systems. For example, one Artificial Intelligence researcher (Kuipers) has been applying these techniques to the domain of renal physiology. Other researchers within the KSL are using similar techniques to build models of cardio-pulmonary physiology.

C. Highlights of Research Progress

C.1 Accomplishments

During the past year we have constructed a second model of the tryptophan operon. The first model we built focused on developing qualitative descriptions of the state variables of the tryptophan operon; this second model focuses on the objects in this biological system, their internal structures, and the processes which modify these objects over time. In addition, we have begun to use this second model as the basis for scientific theory formation. The highlights of this work are summarized below.

C.1.1 Qualitative Modeling and Simulation

Our work in qualitative simulation has been directed towards building a program which embodies a theory of the tryptophan system. The earlier model which we constructed of the system was successful in its ability to predict state variable values for the *trp* system. However, this system lacked flexibility: its fixed network of state variables is valid for only a limited set of experiments. Many experiments involve introducing new state variables, removing old ones, or modifying the interactions between variables. Thus our goal was to build a model of the system which would essentially derive the state variable network used by the earlier system, given a description of what objects were present in an experiment.

In the newer model a gene regulation experiment is described by specifying what objects are present at the start of the experiment and what their properties and relationships are. These objects are represented as instances of prototypical biological objects which are described in a large knowledge base. The modeling system uses a knowledge base of biological processes to detect interactions

between the objects which exist in the current simulation. These interactions can result in the creation of new objects and the establishment of linkages between object state variables, and are used to predict future states of the gene regulation system.

The object knowledge base (KB) is a taxonomic hierarchy of biological objects such as genes, proteins, and chemical binding sites. This KB describes object properties, states, and their decomposition into component objects. The object KB can be viewed as a library of prototypical objects. Modeling of a specific biological experiment begins with a specification of what actual objects (as opposed to prototypes) are present in the experiment. In this context we developed techniques for representing the decomposition of complex objects such as proteins into their components, and for instantiating these prototypical descriptions.

The process knowledge base describes the behaviors of the objects in the trp system. For example, processes encode chemical binding, re-arrangement, and dissociation events which are involved in such biological processes as transcription and biochemical pathways.

When the modeling system is called upon to predict the outcome of an experiment, a process interpreter is responsible for employing processes in the process library to detect interactions between objects in the current experiment. As noted above, these interactions lead to effects such as the creation of new objects in the current experiment, the modification of old objects, and the establishment of linkages between object state variables such as object concentrations.

This model of the trp system has been fully implemented as a working computer program and includes approximately 200 objects and 35 processes. It covers the important components of the trp operon as known in the early 1960s including transcription, translation, and the biosynthetic pathway for tryptophan, and can thus serve as a starting point for the generation of improved theories of the trp operon.

C.1.2 Theory Formation

We have come to view the overall theory formation problem within a machine learning paradigm. Theory formation is considered to be a machine learning problem, which implies that any theory formation program should have two components: a performance element and a learning element. The performance element is the model of the trp operon described above. It contains knowledge of the objects and processes within the trp system, and inference mechanisms for using that knowledge base to predict experimental outcomes. The learning element is used when the performance element makes an incorrect prediction. The learning element must modify the performance element to increase the quality of its predictions.

It appears to be productive to view the theory formation problem as a *design* problem. This allows us to apply knowledge of design within Artificial Intelligence to the problem of theory formation, which is less well understood. Design is a creative activity in which a designer constructs an entity which satisfies a set of constraints. This entity might be an object (e.g., a circuit), or a plan of action (e.g., a robot path plan). The design constraints specify predicates which the design process and the designed entity must satisfy. The entity is constructed from a set of primitives. *Design operators* specify all possible ways in which new primitives may be added to an entity under design. For example, circuits are constructed from transistors and other electronic devices; the design operators describe how these components may be wired together.

AI has approached the problem of design through its central paradigm of search. That is, AI considers the design process to be a kind of search. Search problems

have two important aspects: the space to be searched, and the means by which the search is accomplished. In design problems, the *design space* to be searched is the set of all legal configurations in which the design primitives may be combined to produce designed entities.

The task of theory formation also involves the construction of an entity which satisfies a set of constraints. The entity to be designed is a theory. The primitives from which a theory is designed are the objects and processes within the domain. The constraints on the designed theory include: (a) it must account for, or predict, the observed phenomena, (b) it must be testable, (c) it must be confirmed by the tests, (d) it must conform to a large degree with other established knowledge, (e) it must satisfy certain constraints of form, e.g., it should be as simple as possible.

Our earlier historical study of the discovery of attenuation directed us towards this view of theory formation, and it also provided the design operators needed to make it work. In addition to reconstructing the different theories of the trp operon which the biologists possessed at different times, we also compared consecutive theories to determine the differences between them, which tells us what modifications the biologists applied to existing theories to produce new theories. These theory modification operators are used to design new theories from old. Examples of these operators include postulating the presence of a previously known type of object within an experiment, postulating the presence of a previously *unknown* type of object, and postulating an interaction between existing objects which were not previously thought to interact (a new process).

We have begun implementation of a theory formation program based on the above approach. The inputs to this program are (a) the current theory of the trp operon, (b) a description of an experiment, (c) the outcome of the experiment predicted by the theory, and (d) a description of how the prediction in (c) is incorrect. The output of the program is a set of possible modifications to the theory which cause it to make the correct prediction rather than the prediction in (c).

For example, the initial theory might not predict that a certain chemical X is produced in a given experiment, when in fact this chemical is empirically observed to be present. The theory formation program is implemented as an AI planner; in the above example the planner is given the goal of predicting the presence of X. To achieve this goal it can use a number of different theory formation operators, such as postulating the existence of other chemicals in the experiment, and modifying the processes in the theory such that they would cause X to be produced from the chemicals present in the experiment. The current implementation consists of an agenda-based planner with an incomplete set of these theory formation operators.

D. Publications

1. Bach, R., Friedland, P., Brutlag, D., and Kedes, L.: *MAXIMIZE*, a DNA sequencing strategy advisor. *Nucleic Acids Res.* 10(1):295-304, January, 1982
2. Bach, R., Friedland, P., and Iwasaki, Y.: *Intelligent computational assistance for experiment design*. *Nucleic Acids Res.* 12(1):11-29, January, 1984.
3. Brutlag, D., Clayton, J., Friedland, P. and Kedes, L.: *SEQ: A nucleotide sequence analysis and recombination system*. *Nucleic Acids Res.* 10(1):279-294, January, 1982.
4. Clayton, J. and Kedes, L.: *GEL, a DNA sequencing project management system*. *Nucleic Acids Res.* 10(1):305-321, January, 1982.

- Feitelson, J. and Stefik, M.J.: *A case study of the reasoning in a genetics experiment*. Heuristic Programming Project Report HPP-77-18 (working paper), May, 1977.
5. Friedland, P.: *Knowledge-based experiment design in molecular genetics*. Proc. Sixth IJCAI, August, 1979, pp. 285-287.
 6. Friedland P.: *Knowledge-based experiment design in molecular genetics*. Stanford Computer Science Report STAN-CS-79-760 (Ph.D. thesis), December, 1979.
 7. Friedland, P., Kedes, L. and Brutlag D.: *MOLGEN--Applications of symbolic computation and artificial intelligence to molecular biology*. Proc. Battelle Conference on Genetic Engineering, April, 1981.
 8. Friedland, P.: *Acquisition of procedural knowledge from domain experts*. Proc. Seventh IJCAI, August, 1981, pp. 856-861.
 9. Friedland, P., Kedes, L., Brutlag, D., Iwasaki, Y. and Bach R.: *GENESIS, a knowledge-based genetic engineering simulation system for representation of genetic data and experiment planning*. Nucleic Acids Res. 10(1):323-340, January, 1982.
 10. Friedland, P., and Kedes, L.: *Discovering the secrets of DNA*. Communications of the ACM, 28(11):1164-1186, November, 1985, and IEEE/Computer, 18(11):49:69, November, 1985.
 11. Friedland, P. and Iwasaki Y.: *The concept and implementation of skeletal plans*. Journal of Automated Reasoning, 1(2): 161-208, 1985.
 12. Friedland, P., Armstrong, P., and Kehler, T.: *The role of computers in biotechnology*. BIO\TECHNOLOGY 565-575, September, 1983.
 13. Iwasaki, Y. and Friedland, P.: *SPEX: A second-generation experiment design system*. Proc. of Second National Conference on Artificial Intelligence, August, 1982, pp. 341-344.
 14. Martin, N., Friedland, P., King, J. and Stefik, M.J.: *Knowledge base management for experiment planning in molecular genetics*. Proc. Fifth IJCAI, August, 1977, pp. 882-887.
 15. Meyers, S. and Friedland, P.: *Knowledge-based simulation of regulatory genetics in bacteriophage Lambda*. Nucleic Acids Res. 12(1):1-9, January, 1984.
 16. Stefik, M. and Friedland, P.: *Machine inference for molecular genetics: Methods and applications*. Proc. of NCC, June, 1978.
 17. Stefik, M.J. and Martin N.: *A review of knowledge based problem solving as a basis for a genetics experiment designing system*. Stanford Computer Science Report STAN-CS-77-596, March, 1977.
 18. Stefik, M.: *Inferring DNA structures from segmentation data: A case study*. Artificial Intelligence 11:85-114, December, 1977.
 19. Stefik, M.: *An examination of a frame-structured representation system*. Proc. Sixth IJCAI, August, 1979, pp. 844-852.
 20. Stefik, M.: *Planning with constraints*. Stanford Computer Science Report STAN-CS-80-784 (Ph.D. thesis), March, 1980.

21. Karp, P., and D. Wilkins: *An Analysis of the Deep/Shallow Distinction for Expert Systems*. To be published in *International Journal of Expert Systems*, 1988.
22. Karp, P., and P. Friedland: *Coordinating the Use of Qualitative and Quantitative Knowledge in Declarative Device Modeling*, in *Artificial Intelligence, Simulation, and Modeling* edited by L. Widman and K. Loparo, 1988.
23. Karp, P.: *A Process-Oriented Model of Bacterial Gene Regulation*, Stanford University Knowledge Systems Laboratory Technical Report KSL-88-18.
24. Round, A.: *QSOPS: A Workbench Environment for the Qualitative Simulation of Physical Processes*. Stanford University Knowledge Systems Laboratory Report KSL-87-37, 1987.
25. Karp, P., and P. Friedland, *A Conceptual Reconstruction of the Discovery of Attenuation*. In Preparation.

E. Funding Support

The MOLGEN grant, which has supported the bulk of this research, is titled: MOLGEN: Applications of Artificial Intelligence to Molecular Biology: Research in Theory Formation, Testing, and Modification. This NSF Grant number MCS-8310236, expired on 10/31/86. The Principal Investigators were Edward A. Feigenbaum, Professor of Computer Science and Charles Yanofsky, Professor of Biology. Additional support for this research has been provided by the Defense Advanced Research Projects Agency, under contract N00039-86C-0033.

II. INTERACTIONS WITH THE SUMEX-AIM RESOURCE

SUMEX-AIM continues to serve as the nucleus of our computing resources. The facility has not only provided excellent support for our programming efforts but has served as a major communication link among members of the project. Systems available on SUMEX-AIM such as EMACS, MM, Scribe and BULLETIN BOARD have made possible the project's documentation and communication efforts. The interactive environment of the facility is especially important in this type of project development.

We strongly approve of the network-oriented approach to a programming environment into which SUMEX has evolved. The ability to utilize Lisp workstations for intensive computing while still communicate with all of the other SUMEX resources has been indispensable to our work. We currently have a satisfactory mode of operation where essentially all programming takes place on the workstations and most electronic communications, information sharing, and document preparation takes place within the TOPS-20 environment. The evolution of SUMEX has alleviated most of our previous problems with resource loading and file space. Our current Lisp workstations are not quite fast enough, but we are encouraged by the progress that has been made.

We have taken advantage of the collective expertise on medically-oriented knowledge-based systems of the other SUMEX-AIM projects. In addition to especially close ties with other projects at Stanford, we have greatly benefited by interaction with other projects at yearly meetings and through exchange of working papers and ideas over the system.

III. RESEARCH PLANS

A. *Project Goals And Plans*

Our current work has the following major goals:

1. The abilities of the theory formation system will be extended by implementing additional theory formation operators to allow it to generate new classes of theories.
2. A mechanism for evaluating alternative theories will be constructed. This mechanism will be guide the planner's search towards more plausible theories, and will allow the system to present only the most credible theories it finds to the user.
3. Test the entire approach on the evolving theory of the trp operon regulatory system. Experiment with different initial knowledge bases to see how the discovery process is altered by the availability of new techniques, analogous systems, and so forth.

B. *Justification and Requirements for Continued SUMEX Use*

The MOLGEN project depends heavily on the SUMEX facility. We have already developed several useful tools on the facility and are continuing research toward applying the methods of artificial intelligence to the field of molecular biology. The community of potential users is growing nearly exponentially as researchers from most of the biomedical-medical fields become interested in the technology of recombinant DNA. We believe the MOLGEN work is already important to this growing community and will continue to be important. The evidence for this is an already large list of pilot exo-MOLGEN users on SUMEX.

IV.A.4. ONCOCIN Project

ONCOCIN Project

Edward H. Shortliffe, M.D., Ph.D.
Departments of Medicine and Computer Science
Stanford University

Project Director: Lawrence M. Fagan, M.D., Ph.D.
Department of Medicine
Stanford University

I. SUMMARY OF RESEARCH PROGRAM

A. *Project Rationale*

The ONCOCIN Project is one of many Stanford research programs devoted to the development of knowledge-based expert systems for application to medicine and the allied sciences. The central issue in this work has been to develop a program that can provide advice similar in quality to that given by human experts, and to ensure that the system is easy to use and acceptable to physicians. The work seeks to improve the interactive process, both for the developer of a knowledge-based system, and for the intended end user. In addition, we have emphasized clinical implementation of the developing tool so that we can ascertain the effectiveness of the program's interactive capabilities when it is used by physicians who are caring for patients and are uninvolved in the computer-based research activity.

B. *Medical Relevance and Collaboration*

The lessons learned in building prior production rule systems have allowed us to create a large oncology protocol management system much more rapidly than was the case when we started to build MYCIN. We introduced ONCOCIN for use by Stanford oncologists in May 1981. This would not have been possible without the active collaboration of Stanford oncologists who helped with the construction of the knowledge base and also kept project computer scientists aware of the psychological and logistical issues related to the operation of a busy outpatient clinic.

C. *Highlights of Research Progress*

C.1.A *Background and Overview of Accomplishments*

The ONCOCIN Project is a large interdisciplinary effort that has involved over 35 individuals since the project's inception in July 1979. The work is currently in its ninth year; we summarize here the milestones that have occurred in the research to date:

- *Year 1:* The project began with two programmers (Carli Scott and Miriam Bischoff), a Clinical Specialist (Dr. Bruce Campbell) and students under the direction of Dr. Shortliffe and Dr. Charlotte Jacobs from the Division of Oncology. During the first year of this research (1979-1980), we developed a prototype of the ONCOCIN consultation system, drawing from programs and

capabilities developed for the EMYCIN system-building project. During that year, we also undertook a detailed analysis of the day-to-day activities of the Stanford Oncology Clinic in order to determine how to introduce ONCOCIN with minimal disruption of an operation which is already running smoothly. We also spent much of our time in the first year giving careful consideration to the most appropriate mode of interaction with physicians in order to optimize the chances for ONCOCIN to become a useful and accepted tool in this specialized clinical environment.

- *Year 2:* The following year (1980-1981) we completed the development of a special interface program that responds to commands from a customized keypad. We also encoded the rules for one more chemotherapy protocol (oat cell carcinoma of the lung) and updated the Hodgkin's disease protocols when new versions of the documents were released late in 1980; these exercises demonstrated the generality and flexibility of the representation scheme we had devised. Software protocols were developed for achieving communication between the interface program and the reasoning program, and we coordinated the printing routines needed to produce hard copy flow sheets, patient summaries, and encounter sheets. Finally, lines were installed in the Stanford Oncology Day Care Center, and, beginning in May 1981, eight fellows in oncology began using the system three mornings per week for management of their patients enrolled in lymphoma chemotherapy protocols.
- *Year 3:* During our third year (1981-1982) the results of our early experience with physician users guided both our basic and applied work. We designed and began to collect data for three formal studies to evaluate the impact of ONCOCIN in the clinic. This latter task required special software development to generate special flow sheets and to maintain the records needed for the data analysis. Towards the end of 1982 we also began new research into a *critiquing model* for ONCOCIN that involves "hypothesis assessment" rather than formal advice giving. Finally, in 1982 we began to develop a query system to allow system builders as well as end users to examine the growing complex knowledge base of the program.
- *Year 4:* Our fourth year (1982-1983) saw the departure of Carli Scott, a key figure in the initial design and implementation of ONCOCIN, the promotion of Miriam Bischoff to Chief Programmer, and the arrival of Christopher Lane as our second scientific programmer. At this time we began exploring the possibility of running ONCOCIN on a single-user professional workstation and experimented with different options for data-entry using a "mouse" pointing device. Christopher Lane became an expert on the Xerox workstations that we are using. In addition, since ONCOCIN had grown to such a large program with many different facets, we spent much of our fourth year documenting the system. During that year we also modified the clinic system based upon feedback from the physician-users, made some modifications to the rules for Hodgkin's disease based upon changes to the protocols, and completed several evaluation studies.
- *Year 5:* The project's fifth year (1983-1984) was characterized by growth in the size of our staff (three new full-time staff members and a new oncologist joined the group). The increased size resulted from a DRR grant that permitted us to begin a major effort to rewrite ONCOCIN to run on professional workstations. Dr. Robert Carlson, who had been our Clinical Specialist for the previous two years, was replaced by Dr. Joel Bernstein, while Dr. Carlson assumed a position with the nearby Northern California Oncology Group; this appointment permitted him to continue his affiliation both

with Stanford and with our research group. In August of 1983, Larry Fagan joined the project to take over the duties of the ONCOCIN Project Director while also becoming the Co-Director of the newly formed Medical Information Sciences Program. Dr. Fagan continues to be in charge of the day-to-day efforts of our research. An additional programmer, Jay Ferguson, joined the group in the fall to assist with the effort required to transfer ONCOCIN from SUMEX to the 1108 workstation. A fourth programmer, Joan Differding, joined the staff to work on our protocol acquisition effort (OPAL).

- Year 6: During our sixth year (1984-1985) we further increased the size of our programming staff to help in the major workstation conversion effort. The ONCOCIN and OPAL efforts were greatly facilitated by a successful application for an equipment grant from Xerox Corporation. With a total of 15 Xerox LISP machines now available for our group's research, all full-time programmers have dedicated machines, as do several of the senior graduate students working on the project. Christopher Lane took on full-time responsibility for the integration and maintenance of the group's equipment and associated software. Two of our programming staff moved on to jobs in industry (Bischoff and Ferguson) and three new programmers (David Combs, Cliff Wulfman, and Samson Tu) were hired to fill the void created by their departure and by the reassignment of Christopher Lane.

In addition to funding from DRR for the workstation conversion effort, we have support from the National Library of Medicine which supports our more basic research activities regarding biomedical knowledge representation, knowledge acquisition, therapy planning, and explanation as it relates to the ONCOCIN task domain. We have continued to study the therapy planning process under support from the NLM. This research is led by Dr. Fagan and has concentrated on how to represent the therapy-planning strategies used to decide treatment for patients who run into serious problems while on protocol-described treatment. The physicians who treat these patients often seek out a consultation with the protocol study chairman. Dr. Branimir Sikic, a faculty member from the Stanford University Department of Medicine, and the Study Chairman for the oat cell protocol, collaborated on this project. Janice Rohn joined the ONCOCIN project as data manager and to assist in the knowledge entry process.

- Year 7: The seventh year (1985-86) marked several milestones in our research on workstation-based programming. The OPAL knowledge acquisition system became operational, and several new oncology protocols were entered using this system. David Combs was primarily responsible for creating the operational version of OPAL (based on the initial prototype by Joan Differding Walton). As anticipated, we increased the speed and ease with which protocols can be added to the ONCOCIN knowledge base.

Based on the protocols entered through OPAL, we began experimental testing of the workstation version of ONCOCIN in the Stanford oncology clinic. Clifford Wulfman developed the user interface (based on an initial prototype designed by Christopher Lane). Samson Tu developed the reasoning component (designed originally by Jay Ferguson). Much of their work is built upon an object-oriented system developed for our group by Christopher Lane. We connected the various parts of the system, and demonstrated that we have the capability to run ONCOCIN with the reasoning program and interface program on different machines in the communication network. The current version of the program is currently run on a single workstation, but future versions may take advantage of the multiple machine option. To increase the

speed at which we are able to test protocols entered into ONCOCIN, we developed additional programs to test real and synthetic cases without user interaction; these are then reviewed by our collaborating clinicians.

We also developed a workstation-based program, OPUS, to help clinicians determine which protocols are appropriate for specific patients. OPUS was designed and implemented by Janice Rohn with the assistance of Christopher Lane. We have been using it in the clinic setting since the end of 1985. Thus, in addition to providing an information resource about protocols, the use of a graphically-oriented program provided a way to learn about the software style and hardware used in the workstation version of ONCOCIN.

We discontinued the mainframe version of ONCOCIN, and began using the workstation version exclusively. The performance of the mainframe version of ONCOCIN was documented in two evaluation papers that appeared in clinical journals (see Hickam and Kent's papers).

We continued our basic research in the design of advanced therapy-planning programs: the ONYX project. We developed a model for planning which includes techniques from the fields of artificial intelligence, simulation, and decision analysis. Artificial intelligence techniques are used to create a small number of possible plans given the ideal therapy and the patient's past treatment history. Simulation techniques and decision analysis are used to examine and order the most promising plans. Our goal is to allow ONCOCIN to give advice in a wider range of situations; in particular, the system should be able to recommend plans for patients who have an unusual response to chemotherapy.

During this year, Stephen Rappaport, M.D. joined us as a programmer on the therapy planning research. Clinical expertise for ONCOCIN was provided by Richard Lenon, M.D. and Robert Carlson, M.D.

- *Year 8:* The eight year (1986-87) concentrated on two diverse tasks: 1) scaling up the use of the workstation version of ONCOCIN in the clinic, and 2) generalization of each of the components. The latter task is described in the core research sections of this report(see page 19).

In 1986, we placed the workstation version of ONCOCIN into the Oncology Day Care clinic. This version is a completely different program from the version of ONCOCIN that ran on the DECsystem 20--using protocols entered through the OPAL program, with a new graphical data entry interface, and a revised knowledge representation and reasoning component. One of the Oncology Clinical Fellows (Andy Zelenetz) became responsible for verifying how well our design goals for ONCOCIN had been accomplished. His suggestions have included the addition of key protocols and the ability to have the program used as a data management tool if the complete treatment protocol had not yet been entered into the system. Both of these suggestions were carried out during this year, and the program has achieved wider use in the clinic setting. In addition, laser-printed flowsheets and progress notes have been added to the clinic system.

The process of entering a large number of treatment protocols in a short period of time led to other research topics including: design of an automated system for producing meaningful test cases for each knowledge base, modification of the design and access methods for the time-oriented database, and the development of methods for graphically viewing multiple protocols that are combined into one large knowledge base. These research

efforts will continue into the next year. In addition, some of the treatment regimens developed for the original mainframe version are still in use and can be transferred to the current version of ONCOCIN. As the knowledge base grows, additional mechanisms will be needed for the incremental update and retraction of protocols. Additional changes in the reasoning and interface components of the system are described below.

A new research project related to ONCOCIN was started this last year. We are exploring the use of continuous speech recognition as an alternate entry method for communicating with ONCOCIN. This project requires the connection of speech recognition equipment produced by Speech Systems, Inc. of Tarzana to the ONCOCIN interface module. Christopher Lane has developed a prototype network connection and command interpreter between the speech module (running on a Sun with special hardware added) and the Xerox 1186 computer that runs ONCOCIN. Clifford Wulfman has designed a series of modifications to the ONCOCIN user interface to allow for verbal commands. This work is described in more detail in the core ONCOCIN section.

We continue to collaborate with Andy Zelenetz, Richard Lenon, Robert Carlson, and Charlotte Jacobs on the design and implementation of ONCOCIN in the clinic. Stephen Rappaport has started a residency program to continue his medical education.

- **Year 9:** The majority of our effort this year has been to understand the limitations of the clinic version of ONCOCIN, and to concentrate on the generalization of these techniques to other application areas besides oncology. The majority of this research is thus described as part of the core research discussion on ONCOCIN. Highlights of this year include: (1) development of a general knowledge acquisition tool (PROTEGE) designed to handle skeletal planning applications for clinical trials in any area of medicine, (2) demonstration that the therapy planning and knowledge acquisition tools for ONCOCIN can be closely integrated, and (3) development of a speech input system for ONCOCIN.

As a demonstration of the capabilities of the project to date, we undertook an experiment to see how difficult and time-consuming it is to bring up a new treatment protocol. A summary of a recent colon protocol was downloaded from the PDQ protocol database. Approximately 60% of the knowledge of the protocol summary fit easily into the OPAL high level description. Additional rules were entered using lower level editors. A limited consultation was run after about 4 hours of work. Although this is only one data point, we believe that it validates the generality of the knowledge acquisition and therapy planning approach that we have pursued for nearly a decade. Work continues on extending the knowledge acquisition and therapy planning tools to allow for a higher percentage of concepts that can be entered with the smallest possible amount of low level Lisp changes.

Although we have completed the transfer of ONCOCIN into a stable and useful system on the Xerox Lisp workstations, it is now clear that this type of machine will not provide the type of dissemination hardware we would like to see. There are no planned additions to increase the speed, decrease the cost, or increase the integration capabilities of these workstations. Although there may be other solutions that will allow us to port ONCOCIN directly to alternative hardware platforms, we may need to move away from Xerox workstations and InterLisp language upon which most of our software is based. We are particularly interested in exploring the Mac II hardware recently purchased for the KSL.

C.1.B Review of Research Issues in ONCOCIN and OPAL

Our work to refine the clinic versions of ONCOCIN and OPAL reached a mature stage during this last research year. As our attention has moved to the generalization of these tasks (E-ONCOCIN and PROTEGE) it seems appropriate to describe the range of research issues that we have examined during the development of the ONCOCIN system.

Research Issues in the Development of the ONCOCIN Reasoner and Interviewer

- *Redesign of the reasoning component.* A major impetus for the redesign of the system was to develop more efficient methods to search the knowledge base during the running of a case. We have implemented a reasoning program that uses a discrimination network to process the cancer protocols. This network provides for a compact representation of information which is common to many protocols but does not require the program to consider and then disregard information related to protocols that are irrelevant to a particular patient. We continue to improve portions of the reasoning component that are associated with reasoning over time; e.g., modeling the appropriate timing for ordering tests and identifying the information which needs to be gathered before the next clinic visit. In general, we are concentrating on improving the representation of the knowledge regarding sequences of therapy actions specified by the protocol.

Our experience with adding a large number of protocols has led to the evaluation of the design of the internal structure of the knowledge base (e.g., the way we describe the relationships between chemotherapies, drugs, and treatment visits). We will continue to improve the method for traversing the plan structure in the knowledge base, and consider alternative arrangements for representing the structure of chemotherapy plans. Currently, the knowledge base of treatment guidelines and the patient database are separated. We propose to tie these two structures closer together. Additional work is anticipated on turning ONCOCIN into a critiquing system, where the physician enters their therapy and ONCOCIN provides suggestions about possible alternatives to the entered therapy. Although we have concentrated our review of the ONCOCIN design primarily on the data provided by additional protocols, we know that non-cancer therapy problems may also raise similar issues. The E-ONCOCIN effort is designed to produce a domain-independent therapy planning system that includes the lessons learned from our oncology research. Samson Tu is primarily responsible for continued improvement of the reasoning component of ONCOCIN.

- *Development of a temporal network.* The ability to represent temporal information is a key element of programs that must reason about treatment protocols. The earlier version of the ONCOCIN system did not have an explicit structure for reasoning about time-oriented events. We are experimenting with different configurations of the temporal network, and with the syntax for querying the network. We are also adapting this network so that it can interface with the ONYX therapy-planning systems. This research on temporal reasoning is part of Michael Kahn's Ph.D. thesis. Michael is a student in the Medical Information Sciences Program at University of California at San Francisco.
- *Extensions to the user interface.* We continue to experiment with various configurations of the user interface. Many of the changes have been in response to requests for a more flexible data management environment. We are occasionally faced with data that becomes available corresponding to a

time before the current visit. This can happen if a laboratory result is delayed, or a patient's electronic flowsheet is started in the middle of the treatment. We have added the ability to create new columns of data, and are designing the changes to the temporal processing components of ONCOCIN to allow for data that is inserted out of order. We have also extended the flowsheet to allow for patient specific parameters (e.g., special test results or symptoms) that the physician wishes to follow over time. The flowsheet layouts have been modified to create protocol specific flowsheets, e.g., lymphoma flowsheets have a different configuration than lung cancer flowsheets. The basic structure of the interface has been modified to use object-oriented methods, which allows for more flexible interaction between different components of the flowsheet and the operations performed on the flowsheet.

A continuing area of research concerns how to guide the user to the most appropriate items to enter (based on the needs of the reasoning program) without disrupting the fixed layout of the flowsheet. The mainframe version of ONCOCIN modified the order of items on the flowsheet to extract necessary information from the user. In the workstation version, we have developed a guidance mechanism which alerts the user to items that are needed by the reasoning program. The user is not required to deviate from a preferred order of entry nor required to respond to a question for which no current answer is available. Cliff Wulfman is primarily responsible for improvements to the user interface of ONCOCIN.

- *System support for the reorganization.* The LISP language, which we used to build the first version of ONCOCIN, does not explicitly support basic knowledge manipulation techniques (such as message passing, inheritance techniques, or other object-oriented programming structures). These facilities are available in some commercial products, but none of the existing commercial implementations provide the reliability, speed, size, or special memory-manipulation techniques that are needed for our project. We have therefore developed a "minimal" object-oriented system to meet our specifications. The object system is currently in use by each component of the new version of ONCOCIN and in the software used to connect these components. In addition, all ONCOCIN student projects are now based on this programming environment. Christopher Lane created and is responsible for modifications to the object-oriented system.

Interactive Entry of Chemotherapy Protocols by Oncologists (OPAL)

We continue to refine the software (the OPAL system) that permits physicians who are not computer programmers to enter protocol information on a structured set of forms presented on a graphics display. Most expert systems require tedious entry of the system's knowledge. In many other medical expert systems, each segment of knowledge is transferred from the physician to the programmer, who then enters the knowledge into the expert system. We have taken advantage of the generally well-structured nature of cancer treatment plans to design a knowledge entry program that can be used directly by clinicians. The structure of cancer treatment plans includes:

- choosing among multiple protocols (that may be related to each other);
- describing experimental research arms in each protocol;
- specifying individual drugs and drug combinations;