Form Approved
O.M.B. 68-R0249

DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE
PUBLIC HEALTH SERVICE

GRANT APPLICATION

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

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

1. TITLE OF PROPOSAL *(Do not exceed 53 typewriter spaces)*
S U Medical EXperimental Computer Resource (SUMEX)

2. PRINCIPAL INVESTIGATOR	3. DATES OF ENTIRE PROPOSED PROJECT PERIOD *(This application.*	
2A. NAME *(Last, First, Initial)* Feigenbaum, Edward A.	FROM 08/01/81	THROUGH 07/31/86

2B. TITLE OF POSITION Professor and Chairman Department of Computer Science	4. TOTAL DIRECT COSTS REQUESTED FOR PERIOD IN ITEM 3 $ 6,793,862	5. DIRECT COSTS REQUESTED FOR FIRST 12-MONTH PERIOD $ 1,336,864

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

SUMEX Computer Project – Room TB105
Stanford University Medical Center
Stanford, California 94305

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

Stanford University

2D. DEGREE Ph.D.	**2E. SOCIAL SECURITY NO.**

2F. TELEPHONE DATA	Area Code 415	TELEPHONE NUMBER AND EXTENSION 497-4079

2G. DEPARTMENT, SERVICE, LABORATORY OR EQUIVALENT *(See Instructions)*
Departments of Genetics/Medicine

2H. MAJOR SUBDIVISION *(See Instructions)*
School of Medicine

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

A. [X] NO   B. [ ] YES   Approved: _____
C. [ ] YES – Pending Review   Date _____

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

A. [X] NO   B. [ ] YES – Not previously reported
C. [ ] YES – Previously reported

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

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

Stanford University
Stanford, California 94305
IRS No. 94-1156365
Congressional District No. 12

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

[ ] FEDERAL   [ ] STATE   [ ] LOCAL   [X] OTHER *(Specify)*
Private Non-Profit University

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

K.D. Creighton
Associate Vice President – Controller
Stanford University
Stanford, California 94305

Telephone Number (415) 497-2251

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

Larry J. Lollar
Sponsored Projects Officer
Sponsored Projects Office

Telephone Number(s) (415) 497-2883

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

01   School of Medicine

14. ENTITY NUMBER *(Formerly PHS Account Number)*
IRS No. 94-1156365

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

SIGNATURES *(Signatures required on original copy only. Use ink, "Per" signatures not acceptable)*	A. SIGNATURE OF PERSON NAMED IN ITEM 2A	DATE
	B. SIGNATURE(S) OF PERSON(S) NAMED IN ITEM 10	DATE 5/27/80

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

5/21/80
Date

Edward A. Feigenbaum
Principal Investigator

DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE	LEAVE BLANK
PUBLIC HEALTH SERVICE	
RESEARCH OBJECTIVES	PROJECT NUMBER

NAME AND ADDRESS OF APPLICANT ORGANIZATION

Stanford University, Stanford, California 94305

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

E. Feigenbaum	Principal Investigator	Computer Science
E. Shortliffe	Co—Principal Invest.	Medicine
T. Rindfleisch	Facility Manager	Genetics/Medicine
E. Levinthal	AIM Liaison	Genetics

(See continuation page for additional professional personnel engaged on project.)

TITLE OF PROJECT
Stanford University Medical EXperimental Computer Resource (SUMEX)

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

Stanford University is developing and operating a NATIONAL SHARED COMPUTING RESOURCE in partnership with the NIH Biotechnology Resources Program to explore advanced application of COMPUTER SCIENCE in health research. There are two main objectives of the facility: 1) the managerial, administrative and technical demonstration of a national shared technological resource for health research, and 2) the specific encouragement of application of ARTIFICIAL INTELLIGENCE IN MEDICINE (AIM). Besides the economic advantages of resource sharing made possible by emerging DATA COMMUNICATION technologies, a closer interaction between diverse research efforts is expected to promote a more systematic exchange of research products and ideas. This may be particularly true in applications of computer science. Multilateral community building rather than unilateral service is the project's essential mandate.

The term "artificial intelligence" (AI) is applied to research aimed at increasing the computer's effectiveness as a tool through the emulation of aspects of human SYMBOLIC REASONING and PROBLEM—SOLVING. The field emphasizes the judgmental manipulation of symbolic (non—numeric) representations of knowledge of a task domain for model—building and decision—making. Current applications include programs which assist in inferring chemical structures from spectrographic data, suggesting diagnoses and treatments within various classes of diseases, and modeling aspects of human behavior patterns.

Additional users of the facility will be selected within available resource computer capacity with the help of an AIM Executive Committee and Advisory Group on the basis of reviews of the proposed research. Selection criteria will include general scientific interest and merit, relevance to the AI mission, and community orientation of the collaborator.

LEAVE BLANK

RESEARCH OBJECTIVES (continuation page)

Stanford University Medical Experimental Computer Resource (SUMEX)
Stanford University, Stanford, California 94305

Additional Professional Personnel Engaged on Project:

A. Sweer	System Programmer	Genetics/Medicine
F. Gilmurray	System Programmer	Genetics/Medicine
M. Bizzarri	System Programmer	Computer Science
M. Achenbach	System Programmer	Genetics/Medicine
W. Yeager	System Programmer	Genetics/Medicine
R. Tucker	System Programmer	Genetics/Medicine
B. Buchanan	Adjunct Professor	Computer Science
H.P. Nii	Research Associate	Computer Science
W. van Melle	Research Associate	Computer Science
N. Aiello	Scientific Programmer	Computer Science
N. Veizades	Electronics Engineer	Genetics/Medicine

Biographical Sketches

1   <u>Biographical</u> <u>Sketches</u>

In order to reduce the bulk at the beginning of this already lengthy proposal, we have placed the biographical sketches for all professional personnel contributing to the project in the section starting on page 94.

DETAILED BUDGET FOR FIRST 12-MONTH PERIOD

	FROM	THROUGH
	08/01/81	07/31/82

DESCRIPTION *(Itemize)*		TIME OR EFFORT %/HRS.	AMOUNT REQUESTED *(Omit cents)*		
PERSONNEL			SALARY	FRINGE BENEFITS	TOTAL
NAME	TITLE OF POSITION				
(see next page)	PRINCIPAL INVESTIGATOR				
				462,319	99,645	561,964

CONSULTANT COSTS	None				—

EQUIPMENT			465,000*
Communications, interfaces, test equipment, etc.		10,000	
KI-10 AMPEX core expansion		65,000	
VAX 11-780		250,000	
AIM file server		120,000	
Terminals/displays/printers		20,000	

SUPPLIES			32,000
Computer operations		12,000	
Office supplies		5,000	
Engineering parts		15,000	

TRAVEL	DOMESTIC		6,000
	FOREIGN None		—

PATIENT COSTS None			—
ALTERATIONS AND RENOVATIONS None			—

OTHER EXPENSES			271,900
Equipment maintenance		108,400	
DEC KI-10 (51,000), Calcomp disks/tapes (13,900), DEC 2020 (15,000),			
DEC VAX (10,000), File Server (10,000), DEC PDP-11/GT-40 (4,000),			
Local terminals (4,500)			
Equipment lease		3,000	
Office telephones		7,500	
Local dataphones		10,000	
Software lease and license		6,000	
Technical Services/Repro./Books		4,000	
System and program documentation		3,000	
Network communications		100,000	
SUMEX-AIM collaborative linkages		30,000	

TOTAL DIRECT COST *(Enter on Page 1, Item 5)* ➤ 1,336,864

INDIRECT COST *(See Instructions)*	DATE OF DHEW AGREEMENT:	
	_____ % S&W*	☐ WAIVED
	58 % NTDC August 8, 1979	☐ UNDER NEGOTIATION WITH:
	*IF THIS IS A SPECIAL RATE *(e.g. off-site)*, SO INDICATE.	

Privileged Communication   E. A. Feigenbaum

2.1.2   First Year Personnel Detail

		% Salary
Project Management		
E. Feigenbaum	Principal Investigator	10
E. Shortliffe	Co-Princ Invest	10
T. Rindfleisch	Facility Manager	100
E. Levinthal	AIM Liaison	25
C. Miller	Admin Assistant	100
E. Henderson	Office Assistant	100
D. Vian	Office Assistant	25
System Staff		
A. Sweer	System Programmer	100
F. Gilmurray	System Programmer	100
M. Bizzarri	System Programmer	100
M. Achenbach	Syst Prog/User Cons	100
W. Yeager	Syst Prog/User Cons	100
R. Tucker	Syst Prog/Opns Mgr	100
E. Hedberg	Syst Prog - Stud R.A.	62
J. Clayton	Syst Prog - Stud R.A.	62
Core Research Staff		
B. Buchanan	Adj Professor	10
H. Nii	Research Assoc	60
W. Vanmelle	Research Assoc	50
N. Aiello	Sci Prog	50
P. Cohen	Sci Prog - Stud R.A.	62
D. Smith	Sci Prog - Stud R.A.	62
J. Kunz	Sci Prog - Stud R.A.	62
Electrical Engineering Staff		
N. Veizades	Electronics Engineer	100
E. Schoen	Stud. Electronics Aide	62
Student Syst Prog/Opns Support		
W. Aviles	Syst Prog - Student	50
G. Noga	Syst Prog - Student	50
D. Powers	Syst Prog - Student	50
C. Robinson	Syst Prog - Student	50

Total Salaries	462319
Staff Benefits	99645
********** Total Personnel	561964

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

DESCRIPTION		1ST PERIOD (SAME AS DETAILED BUDGET)	ADDITIONAL YEARS SUPPORT REQUESTED *(This application only)*					
			2ND YEAR	3RD YEAR	4TH YEAR	5TH YEAR	6TH YEAR	7TH YEAR
PERSONNEL COSTS		561,964	621,220	686,694	767,130	848,623	--	--
CONSULTANT COSTS *(Include fees, travel, etc.)*		--	--	--	--	--	--	--
EQUIPMENT (*)		465,000	280,500	416,025	171,576	132,155	--	--
SUPPLIES		32,000	35,200	38,720	42,592	46,851	--	--
TRAVEL	DOMESTIC	6,000	6,600	7,260	7,986	8,785	--	--
	FOREIGN	--	--	--	--	--	--	--
PATIENT COSTS		--	--	--	--	--	--	--
ALTERATIONS AND RENOVATIONS		--	--	--	--	--	--	--
OTHER EXPENSES		271,900	299,995	326,747	346,433	365,906	--	--
TOTAL DIRECT COSTS		1,336,864	1,243,515	1,475,446	1,335,717	1,402,320	--	--

TOTAL FOR ENTIRE PROPOSED PROJECT PERIOD *(Enter on Page 1, Item 4)* ⟶	$	6,793,862

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

(*) Equipment Purchase items are not included in the Net Total Direct Cost base used to compute Indirect Costs.

(see continuation pages for budget justification)

2.3   Budget Explanation and Justification

The following paragraphs explain in detail our budget plan over the proposed 5-year grant term.   Indirect costs are not shown in the budget and will be computed separately on the basis of Net Total Direct Costs (Total Direct Costs less funds for Equipment Purchase).   In the most recent agreement between Stanford and the DHEW dated August 8, 1979, the indirect cost rate is 58%.

Personnel

The proposed personnel budget is based on the current staffing for resource management, development, and operations with the addition of a system programmer and an engineering aide to support planned new hardware and software development work.   Individual salary figures are not included in the "first year budget detail" plan but have been submitted separately to NIH in confidence.   The salary estimates reflect current actual rates and include anticipated increases averaging 10% annually based on recent experience with inflation.   Staff benefits are computed using rates currently projected by Stanford University: 21.0% for 8/81, 21.6% for 9/81-8/82, 22.2% for 9/82-8/83, 22.8% for 9/83-8/84, 24.8% for 9/84-8/85, and 25.4% for 9/85-8/86.

Project Management and Technical Direction:

Prof. Feigenbaum is budgeted at 10% as project principal investigator, Prof. Shortliffe at 10% as co-principal investigator for medical liaison (*), Mr. Rindfleisch at 100% is responsible for facility implementation and management, Dr. Levinthal at 25% is responsible for liaison with the national AIM community and the AIM management committees, and Ms. Miller and Ms. Henderson at 100% each provide project administrative and office assistance for SUMEX and community affairs.

System programming:

The programming staff, while sharing a substantial joint responsibility for system development/maintenance, user assistance, subsystem and utility program development, and operational support, have specific areas of responsibility as follows.   Messrs. Sweer and Gilmurray, and Bizzarri (100% each) share responsibility for monitor and system support.   These duties include, for example, on-going development work for new machine integration into the facility, Ethernet implementation, performance analysis and improvement, system communications support, special device drivers and diagnostics, scheduler controls, and system maintenance.   They also share responsibility for system software such as

----

(*) No salary is shown for Dr. Shortliffe for the first 3 years because he is supported by an NLM Research Career Development Award through 6/84.   In order to assist his work on the project, we budget 25% support for D. Vian, his office assistant

EXECutive programs, languages, and other general utilities. Mr. Hedberg is
a student system programmer who has been working with the project for
several years and will continue to work on EXEC developments, network
interface software, and software compatibility under supervision of the
system staff.

System maintenance and operations:

Mr. Tucker (100%) is responsible for our network liaison, operations
utility program development and maintenance, and overseeing system
operations and backup. He is assisted in providing file system
archive/restore service and backup dumps as well as system utility
programming support by the four undergraduate students (currently Messrs.
Aviles, Noga, Powers, and Robinson).

User support:

The user support staff includes Mr. Michael Achenbach (100%), Mr.
William Yeager (100%), and a student research assistant, Ms. Jan Clayton.
Messrs. Achenbach and Yeager will share responsibility for subsystem
maintenance and user consulting as well as assisting with software to
integrate planned new hardware. Mr. Achenbach also assists in interfacing
user program packages into the system (e.g., DENDRAL, MYCIN), assuring
appropriate documentation and assisting with initial user contacts. Mr.
Yeager serves as the primary contact for user consultation, answering many
questions himself and referring others to the appropriate staff members
expert in particular areas. Mr. Yeager will also continue development of
inter-user communication facilities. Ms. Clayton will be responsible for
updating system documentation and developing more effective tools for users
to access available documentation.

AI Core Research:

We budget partial support for specific members of the Heuristic
Programming Project for core research work to explore basic AI issues
relating to biomedical applications and to develop and generalize AI
software tools important to the entire SUMEX-AIM community. Complementary
support for related work within the HPP is received from other sources such
as ARPA and NSF. Prof. Buchanan (10%) will provide technical direction for
staff and students working on proposed core research efforts. Ms. Nii
(60%) and Dr. Vanmelle (50%) will lead the AGE and EMYCIN efforts
respectively. Ms. Aiello (50%) will provide programming support and the
graduate research assistants, Messrs Cohen, Smith, and Kunz will work on
thesis topics related to particular core research goals.

Electronics support:

Finally we budget Mr. Veizades (100%) and a student engineering aide for hardware engineering and maintenance. They are responsible for designing needed special purpose hardware (e.g., communications equipment, intermachine network hardware, and Ethernet interfaces), integrating new hardware into the facility, and maintaining facility equipment.

Consultant

We do not now plan any consulting support during the follow-on grant period.

Equipment

The "Equipment" budget covers only equipment purchases. Lease arrangements for collaborator terminal and communications support as well as maintenance contracts are discussed under "Other".

Minor Equipment:

$10,000 per year is allocated for minor equipment purchases including communications equipment, Ethernet interfaces, and test equipment. This budget is increased by 5% per year to accommodate inflation.

Major Equipment:

Following are budget estimates for the major equipment acquisitions planned. The prices quoted are best current estimates. Over the 5-year term of the grant prices will certainly change and alternate vendor options may become available for some subsystems. We will carefully review each purchase with BRP to achieve the most advantage in terms of technical and cost effectiveness.

yr 1 - Add 256K words of core to the existing KI-10 AMPEX memory to reduce page swapping overhead. This will cost $65,000 based on a quote from AMPEX for the memory modules and control logic to augment the existing ARM-10LX cabinet.

- Buy a VAX 11/780 with 2M bytes of memory, floating point accelerator, 1 RP-06 disk drive, 1 TE-16 tape drive, and 1 DZ-11 line group at $250,000 based on a current price quotation including tax. This machine will be used to provide large address space INTERLISP facilities, to experiment with AI program export, to support development of VAX system software for the community, and to alleviate congestion in the Stanford 40% of the SUMEX resource. This system has minimal memory for this initial integration work and will be expanded in year 2.

- Buy a bare PDP-11/34 processor with 64K of memory ($18,000), 2 Trident 300 Mbyte disk drives with controller ($49,000), and 2 STC 6250 BPI magnetic tape drives with controller ($53,000) to develop a community file server. This file server will be coupled to SUMEX host machines via the high speed Ethernet. This will minimize the need for redundant large file systems on each host and alleviate the file storage limitations of the AIM community.

- $20,000 is allocated for a "Stanford University Network" bit-mapped display terminal station ($10,000) and a Canon laser printer for high quality hardcopy output ($10,000).

yr 2 - Add 2M bytes of memory to the VAX purchased in year 1 ($70,000).

- Add 630M bytes to the file server purchased in year 1 ($40,000). This will include 2 300 Mbyte drives which will fill the controller.

- Buy 5 single-user "professional workstations" (PWS) ($160,000 -- $30,000 each plus tax). This price is based on the projected cost of the Zenith-MIT NU system or its equivalent. These machines will be used to develop and experiment with user-dedicated machines for AI program development, export, and human interface enhancements. These machines will be distributed within the Stanford community initially to facilitate development and will be coupled by Ethernet with the main resource.

yr 3 - Add a second VAX 11/780 with 4 Mbytes memory, 1 RP-06 disk drive, 1 TE-16 tape drive, floating point accelerator, and 1 DZ-11 line group ($320,000) for general community support with large address space INTERLISP. This machine will be managed for program testing in a way similar to the existing 2020.

- Add 2 PWS systems ($65,000) to be distributed within the AIM community under Executive Committee control.

- $20,000 is allocated for an additional "Stanford University Network" bit-mapped display terminals ($10,000) and a Canon laser printer for high quality hardcopy output ($10,000) for the anticipated growing and distributed community of local users.

yr 4 - Add 3 PWS systems ($100,000) to be distributed within the AIM community under Executive Committee control.

- Add 630M bytes to the central file server to meet expected growth in community file storage needs. This will include a second controller with two drives ($60,000)

yr 5 - Add 3 PWS systems ($100,000) to be distributed within the AIM community under Executive Committee control.

- $20,000 is allocated for an additional "Stanford University
  Network" bit-mapped display terminals ($10,000) and a Canon laser
  printer for high quality hardcopy output ($10,000) for the
  anticipated growing and distributed community of local users.

Supplies

The computer supplies budget is an extension of our recent operating
experience with the SUMEX-AIM facility and expected increases for the new
machines. We estimate $12,000 for the first year covering paper, ribbons,
tapes, disk packs, labels, and other supplies. We budget a 10% per year
escalation of these costs. Office supplies are budgeted at $5,000 per year
also based on past experience and are increased 10% per year. Engineering
supplies cover needed parts and spares for interfacing and integrating new
equipment and for maintaining in-house equipment. We budget $15,000 per
year for this purpose with an annual inflation factor of 10%.

Travel

The travel budget covers travel to technical meetings, management
committee meetings, and AIM workshop meetings as well as travel to assist
user groups get started on SUMEX as needed. We budget for 4 east coast
trips ($800 each), 3 midwest trips ($600 each), and 4 west coast trips
($250 each). Future years are inflated by 10% per year.

Other

Equipment Maintenance:

We budget for facility equipment maintenance based on our past
experience with DEC and other vendors. We expect to retain our favorable
cooperative maintenance arrangements with DEC for the KI-10 and 2020
systems and to add appropriate vendor contracts for the other equipment
(VAX's, file server, Professional workstations, etc.) as acquired. We
spend substantial staff effort in maintaining equipment to minimize costs
in contracts and "time and materials" to outside vendors. We continue to
investigate alternatives for maintenance: either in-house or from another
vendor. So far we have not been able to project enough cost savings or
improved service to justify a change. With costs continuously rising, we
will periodically re-evaluate alternatives to achieve the most cost
effective maintenance service for the resource. We have budgeted a 5% per
year inflation for maintenance costs.

Equipment Lease:

We budget $3,000 per year for equipment lease related to on-going collaborative linkages to SUMEX. $2,000 per year is allocated for continued lease of a communication line between the SUMEX machine room and the SECS facilities at the University of California at Santa Cruz. $1,000 per year is for a line to Prof. Langridge's group at UC San Francisco. These lines were approved by the AIM Executive Committee.

Telephone Services:

We budget $7,500 per year for staff office and home terminal telephones and $10,000 per year to cover dataphone services for local Stanford community dialup ports on the SUMEX computer. These estimates are based on the current configuration of lines and expected growth for planned new equipment. We periodically review these arrangements to maintain satisfactory service at minimum cost.

Software Lease:

We budget $6,000 per year for software lease costs. These funds are used to maintain our license rights to and updates for such software as DEC monitors, language and utility products, SITBOL, STP, SPSS, SIMULA, etc. as well as additional packages the community may require.

Services and Documentation:

$4,000 per year is budgeted for books, publications, technical services, and reproduction based on previous experience. $3,000 per year is budgeted for providing to users up-to-date documentation for system and subsystem usage. Substantial efforts continue to upgrade documentation for the user community.

Communications support:

We budget a total of $100,000 per year for network services starting in year 1 and increased by 5% per year. Of this amount, $75,000 is allocated based on current experience for TYMNET services (including network interface, maintenance, and usage costs) projected to accommodate increased usage for the new equipment. In past years, these funds have been distributed directly from NIH/BRP through NLM contracts with TYMNET. This may still prove to be the most cost-effective approach and we will work closely with NIH/BRP to secure these critical services at the lowest cost.

The remaining $25,000 is budgeted as a contingency to experiment with other networks or communications media to support AIM work if justified by community needs and technological developments or to retain our highly beneficial ARPANET connection. A growing number of the AIM community

members with local machines have expressed the need for a means to transfer files with SUMEX. This need will increase with more distributed AIM computing resources. Since TYMNET is not currently moving to provide this kind of service, further experimentation with TELENET or other vendors may be warranted.

At present SUMEX-AIM ARPANET costs are being borne by ARPA-IPTO as part of the Stanford Heuristic Programming Project contract. We have no information that this relationship will change (we do get frequent inquiries from ARPA about its status however). The $25,000 contingency may be needed to cover part of these costs should ARPA/DCA policies changes.

Collaborative Linkages:

We budget $30,000 per year for collaborative linkage needs. These funds will be available for terminals, lines, and other facilities to enable more effective inter-group collaborations and contacts with medical scientists. These funds have been very effective in the past in assisting new projects get connected to available computing resources within the AIM community pending grant support of their research. These funds are allocated in close cooperation with the AIM Executive Committee and BRP. We budget a 5% annual increase for this collaborative linkage support.

II. Research Plan

This is an application for renewal of a grant supporting the Stanford University Medical EXperimental computer research resource for applications of Artificial Intelligence in Medicine (SUMEX-AIM). We have attempted to keep this proposal as brief as possible and to place detailed background information in appendices. However, we felt obliged to exceed some of the page limitations stipulated in the NIH guidelines for a several reasons:

1) the computer science discipline of artificial intelligence is relatively new and its intersection with and significance to medicine requires more explanation than more traditional areas of biomedical research.

2) the SUMEX-AIM resource encompasses a national community of more than 20 research projects pursuing diverse applications areas. In order to illustrate the scope of the community and to provide the scientific basis for continued support of SUMEX as a resource, the objectives of these projects must be presented. We also include a brief description of the important operational base of the resource that may be unfamiliar to some reviewers.

3) this application is for a 5-year renewal term. Many of the core and collaborative research efforts are aimed at long term goals to assist biomedical researchers and clinicians in information management, analysis, and decision making. In order to provide a more efficient research environment, avoiding the overhead of additional proposal preparations and reviews on time scales shorter than expected result horizons, we hope to describe our goals in sufficient detail to justify the 5-year award period.

3   Introduction and Aims

3.1   Overview of Objectives and Rationale

The SUMEX-AIM ("SUMEX") project is a national computer resource with a dual mission: a) the promotion of applications of computer science research in artificial intelligence (AI) to biological and medical problems and b) the demonstration of computer resource sharing within a national community of health research projects. The SUMEX-AIM resource is located physically in the Stanford University Medical School and serves as a nucleus for a community of medical AI projects at universities around the country. SUMEX provides computing facilities tuned to the needs of AI research and communication tools to facilitate remote access, inter- and intra-group contacts, and the demonstration of developing computer programs to biomedical research collaborators.

In the body of this proposal, we offer definitions and explanations of these efforts at several levels of detail to meet the needs of reviewers from various perspectives. For this overview, we give only a brief definition of AI and a summary of the background, present status, and expectations of our research for the requested term of the renewal, the five years beginning August 1, 1981.

3.1.1   Definitions of Artificial Intelligence

Artificial Intelligence research is that part of Computer Science concerned with symbol manipulation processes that produce intelligent action [1 - 7]. By "intelligent action" is meant an act or decision that is goal-oriented, is arrived at by an understandable chain of symbolic analysis and reasoning steps, and utilizes knowledge of the world to inform and guide the reasoning.

Placing AI in Computer Science

A simplified view relates AI research with the rest of computer science. The manner of use of computers by people to accomplish tasks can be "one-dimensionalized" into a spectrum representing the nature of the instructions that must be given the computer to do its job; call it the WHAT-TO-HOW spectrum. At the HOW extreme of the spectrum, the user supplies his intelligence to instruct the machine precisely HOW to do his job, step-by-step. Progress in computer science may be seen as steps away from that extreme "HOW" point on the spectrum: the familiar panoply of assembly languages, subroutine libraries, compilers, extensible languages, etc. illustrate this trend.

At the other extreme of the spectrum, the user describes WHAT he wishes the computer to do for him to solve a problem. He wants to communicate WHAT is to be done without having to lay out in detail all

necessary subgoals for adequate performance yet with a reasonable assurance
that he is addressing an intelligent agent that is using knowledge of his
world to understand his intent, complain or fill in his vagueness, make
specific his abstractions, correct his errors, discover appropriate
subgoals, and ultimately translate WHAT he wants done into detailed
processing steps that define HOW it shall be done by a real computer. The
user wants to provide this specification of WHAT to do in a language that
is comfortable to him and the problem domain (perhaps English) and via
communication modes that are convenient for him (including perhaps speech
or pictures). The research activity aimed at creating computer programs
that act as "intelligent agents" near the WHAT end of the WHAT-TO-HOW
spectrum can be viewed as a long-range goal of AI research.

Expert Systems and Applications

The national SUMEX-AIM resource is an outgrowth of a long,
interdisciplinary line of artificial intelligence research at Stanford
concerned with the development of concepts and techniques for building
"expert systems" [1]. An "expert system" is an intelligent computer
program that uses knowledge and inference procedures to solve problems that
are difficult enough to require significant human expertise for their
solution. For some fields of work, the knowledge necessary to perform at
such a level, plus the inference procedures used, can be thought of as a
model of the expertise of the expert practitioners of that field.

The knowledge of an expert system consists of facts and heuristics.
The "facts" constitute a body of information that is widely shared,
publicly available, and generally agreed upon by experts in a field. The
"heuristics" are the mostly-private, little-discussed rules of good
judgment (rules of plausible reasoning, rules of good guessing) that
characterize expert-level decision making in the field. The performance
level of an expert system is primarily a function of the size and quality
of the knowledge base that it possesses.

Currently authorized projects in the SUMEX community are concerned in
some way with the application of AI to biomedical research (*). The
tangible objective of this approach is the development of computer programs
that will be more general and effective consultative tools for the
clinician and medical scientist. There have already been promising results
in areas such as chemical structure elucidation and synthesis, diagnostic
consultation, and modeling of psychological processes.

Needless to say, much is yet to be learned in the process of
fashioning a coherent scientific discipline out of the assemblage of
personal intuitions, mathematical procedures, and emerging theoretical
structure comprising artificial intelligence research. State-of-the-art
programs are far more narrowly specialized and inflexible than the
corresponding aspects of human intelligence they emulate; however, in

---

(*) Brief abstracts of the various projects can be found in Appendix
A on page 331 and more detailed progress summaries in Section 9 on page
135.

special domains they may be of comparable or greater power, e.g., in the
solution of formal problems in organic chemistry.

3.1.2   Resource Sharing

An equally important function of the SUMEX-AIM resource is an
exploration of the use of computer communications as a means for
interactions and sharing between geographically remote research groups
engaged in biomedical computer science research. This facet of scientific
interaction is becoming increasingly important with the explosion of
complex information sources and the regional specialization of groups and
facilities that might be shared by remote researchers [8]. We expect an
even greater decentralization of computing resources in the coming years
with the emerging VLSI (*) technology in microelectronics and a
correspondingly greater role for digital communications.

Our community building effort is based upon the current state of
computer communications technology. While far from perfected, these
developing capabilities offer highly desirable latitude for collaborative
linkages, both within a given research project and among them. A number of
the active projects on SUMEX are based upon the collaboration of computer
and medical scientists at geographically separate institutions; separate
both from each other and from the computer resource. The network
experiment also enables diverse projects to interact more directly and to
facilitate selective demonstrations of available programs to physicians,
scientists, and students.

We have actively encouraged the development of additional affiliated
computing resources within the AIM community. Since 1977, the facility at
Rutgers University has allocated a portion of its capacity for national AIM
projects and our network connections to Rutgers and common facilities for
user terminals have been indispensable for effective interchanges between
community members, workshop coordinations, and software sharing.

Even in their current developing state, communication facilities
enable effective access to the specialized SUMEX computing environment from
a great many areas of the United States and to a more limited extent from
Canada, Europe, Australia, and other international locations.

3.2   SUMEX-AIM Background

Beginning in the mid-1960's with DENDRAL (**), a project focused on
applications of artificial intelligence to problems of biomolecular

----

(*) Very Large Scale Integration

(**) Much of the early DENDRAL computation work was done on the ACME
IBM 360/50 interactive computing resource at Stanford, which was funded by
the NIH Biotechnology Resources Program between 1965 and 1973.

structure characterization, the Stanford Heuristic Programming Project has
pioneered in expert systems research with funding support from NIH, ARPA,
NSF, and NASA. Since 1973, SUMEX-AIM has developed as a national resource
for applying these techniques to a broad range of biomedical research
problems.

Funding of the SUMEX-AIM resource from the NIH Biotechnology
Resources Program (BRP) began in December 1973 for a five year period.
Prof. Joshua Lederberg was Principal Investigator and Prof. Edward A.
Feigenbaum was co-Principal Investigator. The major hardware was delivered
and accepted in April 1974, and the system became operational for users
during the summer of 1974. In 1977, we applied for a five-year renewal
grant to continue our national research effort. We received a
recommendation for approval of the five year period from the study section
but this was reduced to three years following Professor Lederberg's
decision in early 1978 to accept the presidency of The Rockefeller
University. The principal investigator role passed easily to Prof.
Feigenbaum, Chairman of the Stanford Computer Science Department, based
upon his long-time involvement with the project and close collaboration
with Prof. Lederberg. The highly interdisciplinary spirit of SUMEX has
been retained with very close ties to the Stanford Medical School through
Drs. E. H. Shortliffe (current co-Principal Investigator of SUMEX) and S.
N. Cohen.

Although six years is hardly long enough for a conclusive
determination of the success of the SUMEX-AIM model, we can fairly take
pride in the diligence and technical competence with which we have
responded to the community responsibilities mandated by the terms of our
grant. An important element in satisfying those responsibilities was the
establishment of a mutually satisfactory management structure, on which we
report in further detail later (see Appendix E on page 383). Good will and
common purpose are of course the indispensable ingredients for an effective
community resource, and we are grateful to have been able to offer this
service in a congenial framework, and at the same time to be able to
support our local computing research needs.

The present renewal application is therefore written from a
perspective of having built a substantial community of active biomedical AI
research projects and having just begun the new phase of our research to
integrate and exploit emerging computer technologies that will have a
profound effect on the development and export of practical medical AI
programs. Beginning with 5 projects in 1973, the AIM community grew to 11
major projects at our renewal in 1978 and currently numbers 17 fully
authorized projects plus a group of 8 pilot efforts. In addition to the
Rutgers Computers in Biomedicine project, two of the formal projects and
one of the pilots do their computing using the portion of the Rutgers
University facility allocated to AIM community users. As discussed in the
sections describing the individual projects (see Section 9 on page 135),
many of the computer programs under development by these groups are
maturing into tools increasingly useful to the respective research
communities. The demand for production-level use of these programs has
surpassed the capacity of the present SUMEX facility and has raised
important issues of how such software systems can be optimized for
production environments, exported, and maintained.

3.3.1   Resource Operations

1) Maintain the vitality of the AIM community. We will continue to
    encourage and explore new applications of AI to biomedical research
    and improve mechanisms for inter- and intra-group collaborations and
    communications. While AI is our defining theme, we may entertain
    exceptional applications justified by some other unique feature of
    SUMEX-AIM essential for important biomedical research. To minimize
    administrative barriers to the community-oriented goals of SUMEX-AIM
    and to direct our resources toward purely scientific goals, we plan
    to retain the current user funding arrangements for projects working
    on SUMEX facilities. User projects will fund their own manpower and
    local needs; will actively contribute their special expertise to the
    SUMEX-AIM community; and will receive an allocation of computing
    resources under the control of the AIM management committees. There
    will be no "fee for service" charges for community members. We will
    also continue to exploit community expertise and sharing in software
    development; and to facilitate more effective information sharing
    among projects.

2) Continue to provide effective computational support for AIM
    community goals. Our efforts will be to extend the support for
    artificial intelligence research and new applications work; to
    develop new computational tools to support more mature projects; and
    to facilitate testing and research dissemination of nearly
    operational programs. We will continue to operate and develop the
    existing KI-10/2020 facility as the nucleus of the resource. We
    will acquire additional equipment to meet developing community needs
    for more capacity, larger program address spaces, and improved
    interactive facilities. New computing hardware technologies
    becoming available now and in the next few years will play a key
    role in these developments and we expect to take the lead in this
    community for adapting these new tools to biomedical AI needs. We
    plan the phased purchase of two VAX computers to provide increased
    computing capacity and to support large address space LISP
    development, a 2000M byte file server to meet file storage needs,
    and a number of single-user "professional workstations" to
    experiment with improved human interfaces and AI program
    dissemination.

3) Provide effective and geographically accessible communication
    facilities to the SUMEX-AIM community for effective remote
    collaborations, communications among distributed computing nodes,
    and experimental testing of AI programs. We will retain the current
    ARPANET and TYMNET connections for at least the near term and will
    actively explore other advantageous connections to new
    communications networks and to dedicated links.

3.3.2   Training and Education

        Our goals during the follow-on period for assisting new and
established users of the SUMEX-AIM resource are a continuation of those
adopted for the previous grant term. Collaborating projects are
responsible for the development and dissemination of their own AI programs.
The SUMEX resource will provide community-wide support and will work to
make resource goals and AI programs known and available to appropriate
medical scientists. Specific aims include:

   1)   Provide documentation and assistance to interface users to resource
        facilities and programs. We will continue to exploit particular
        areas of expertise within the community for developing pilot efforts
        in new application areas.

   2)   Continue to allocate "collaborative linkage" funds to qualifying new
        and pilot projects to provide for communications and terminal
        support pending formal approval and funding of their projects.
        These funds are allocated in cooperation with the AIM Executive
        Committee reviews of prospective user projects.

   3)   Continue to support workshop activities including collaboration with
        the Rutgers Computers in Biomedicine resource on the AIM community
        workshop and with individual projects for more specialized workshops
        covering specific application areas or program dissemination.

3.3.3   Core Research

        Our core research efforts will continue to emphasize basic research
on AI techniques applicable to biomedical problems and the generalization
and documentation of tools to facilitate and broaden application areas.

        SUMEX core research funding is complementary to similar funding from
other agencies and contributes to the long-standing interdisciplinary
effort at Stanford in basic AI research and expert system design. We
expect this work to provide the underpinnings for increasingly effective
consultative programs in medicine and for more practical adaptations of
this work within emerging microelectronic technologies. Specific aims
include:

   1)   Continue to explore basic artificial intelligence issues for
        knowledge acquisition, representation, and utilization; reasoning in
        the presence of uncertainty; strategy planning; and explanations of
        reasoning pathways with particular emphasis on biomedical
        applications.

   2)   Support community efforts to organize and generalize AI tools that
        have been developed in the context of individual application
        projects. This will include work to organize the present state-of-
        the-art in AI techniques through the AI Handbook effort and the

development of practical software packages (e.g., AGE, EMYCIN,
UNITS, and EXPERT) for the acquisition, representation, and
utilization of knowledge in AI programs. The objective is to evolve
a body of software tools that can be used to more efficaciously
build future knowledge-based systems and explore other biomedical AI
applications. The details of these are given in Section 6.3.

4   Significance

What is the significance of the artificial intelligence research and knowledge engineering work for which SUMEX is a resource? And what is the significance of SUMEX for achieving the goals of the enterprise?

In this section, we first sketch, in an abstract way, the significance of the scientific work. We then probe more deeply examining medicine, biochemistry, and psychology. Finally, we look at SUMEX's facilitative role, particularly in the light of the microelectronic revolution; and conclude with a discussion of the more general aspects of SUMEX's scientific role in enhancing scientific communication and knowledge.

A Brief Recapitulation

Artificial Intelligence research and its applications-oriented twin, Knowledge Engineering, are those parts of Computer Science that are concerned with the representation of symbolic knowledge for computer use; and the construction of programs for symbolic inference that can make use of the knowledge to achieve intelligent action. Examples of such actions include finding problem solutions, forming hypotheses, offering advice, inferring diagnoses, recommending therapeutic steps, and so on. The knowledge that must be used is a combination of factual knowledge and heuristic knowledge. The latter is especially hard to obtain and represent since the experts providing it are mostly unaware of the heuristic knowledge they are using.

Managing the Growth of Knowledge

Medical and scientific communities currently face many problems relating to the rapid cumulation of knowledge, for example:

- codification of theoretical and heuristic knowledge

- effective use of the wealth of information implicitly available in textbooks, journal articles and from practitioners

- dissemination of that knowledge beyond the intellectual centers where it is collected

- customizing the presentation of that knowledge to individual practitioners as well as customizing the application of the information to individual cases

These needs are widely recognized. In addition, computers are recognized as the most hopeful technology to overcome the problems. While recognizing the value of mathematical modeling, statistical classification, decision theory and other techniques, we believe that effective use of those methods depends on using them in conjunction with less formal knowledge, including contextual and strategic knowledge.

Artificial intelligence offers advantages for representing information and using it that will allow physicians and scientists to use computers as intelligent assistants. In this way we envision a significant extension to the decision making powers of individual practitioners without reducing the significance of the individuals.

More specifically...AI in the service of Medicine

Although computing technology is playing an increasingly important role in medicine, systems designed to advise physicians on diagnosis or therapy selection have received poor clinical acceptance. Despite diverse research efforts, and a literature on computer-aided diagnosis that has numbered at least 1000 references in the last 20 years, clinical consultation programs have seldom been used other than in experimental environments.

The reasons for attempting to develop such systems are self-evident. Growth in medical knowledge has far surpassed the ability of the single practitioner to master it all, and the computer's superior information processing capacity thereby offers a natural appeal. Furthermore, the reasoning processes of medical experts are poorly understood; attempts to model expert decision making necessarily require a degree of introspection and a structured experimentation that may in turn improve the quality of the physician's own clinical decisions, making them more reproducible and defensible. New insights that result may also allow us more adequately to teach medical students and house staff the techniques for reaching good decisions, rather than merely to offer a collection of facts which they must independently learn to utilize coherently.

In recent years observers have begun to analyze the reasons for poor acceptance of the systems that have sprung from such research, and some have argued that the problems have tended to lie not only with the decision-making performance of such programs but also with system design features that have failed to appreciate the physician's viewpoint or have made the interactive process unappealing. To correct,these deficiencies future systems must be fast, easy to use, and congenial. They must address important clinical problems with which physicians recognize they need assistance. But perhaps most important, in order to stress the primary physician's role as ultimate decision maker, they must be able to explain what they are doing, not through quotations of statistical theory but in terms of a line of reasoning that is familiar and similar to the kind of justification a clinician might expect from a human consultant. Explanation capabilities help the physician using the program decide whether to follow its advice; they thereby emphasize the computer's function as a helpful tool that is intended to complement rather than replace the primary physician's own decision-making powers.

Because of considerations such as these, the last decade has witnessed the development of new approaches to computer-based medical decision making. Of particular significance is research directed at the encoding and utilization of experts' judgmental knowledge -- the kind of practical experience which underlies the daily practice of medicine and is

Significance

far-removed from the mathematical approaches of formal decision analysis.
Artificial Intelligence is a particularly relevant computer science
subfield because of its emphasis on symbolic reasoning capabilities rather
than numeric computations. The AIM community's promising research into
medical symbolic reasoning represents more than the application of well-
established computing techniques. Although the approaches are young and
experimental, significant accomplishments in codifying medical knowledge
and modeling clinical reasoning have already been achieved. Additional
investigation, in artificial intelligence and in related computer science
subfields, will further facilitate the development of useful, congenial,
high-performance consultation systems. These systems will improve when we
know better how to manage such problems as (1) understanding the psychology
of medical reasoning as practiced by specialists, (2) automated
interpretation of written and spoken natural language, (3) acquisition and
representation of knowledge obtained from collaborating experts, (4)
encoding and utilization of time relationships central to many disease
processes, and (5) mechanisms for representing and measuring inexact
reasoning.

AI in the service of Biochemistry: why SUMEX?

Consider three major projects engaged in research in structural
biochemistry:

1) DENDRAL, computer-assisted elucidation of molecular structure,
   including stereochemistry, with applications in the areas of natural
   products, bio-active compounds and conformational analysis

2) MOLGEN, investigations of experiment planning in molecular genetics,
   including structural studies of large biomolecules with emphasis on
   sequencing of nucleic acids

3) SECS, computer simulation and evaluation of chemical synthesis

In each case, a new type of computational assistance is being made
available to a significant modern area of scientific research. Though in
the past each field has made some use of the numeric and searching
capabilities of computers, the use of advanced methods for symbolic
manipulation, representation of knowledge, and inference is new, currently
significant, and holds great promise in future development.

Over the past several years all three projects have matured to the
point where specific programs are being disseminated to the scientific
community via the mechanisms of outside access to SUMEX or direct program
export to other laboratories. Each project is currently engaged in studies
pointed toward both application of existing programs to real biochemical
problems and research into new computer-based tools for future
applications. The SUMEX resource provides a focal point for building a
collaborative community with common interests in particular programs. The
resource provides the computational capacity for new developments and a
medium for communication for discussions of successes, and failures, aimed
at improving application programs.

The rapid development of these programs, to the point of sharing the programs with a community of investigators, is due to several factors. These factors are important in understanding the special significance of the SUMEX resource and the role it plays in continued development and dissemination of the programs. All three projects share an important underlying thread, and that is the concept of a molecular structure. Even though the three projects deal with computer representations of molecular structures at varying levels of specificity, the fact that there are formal, precise descriptions of structure available greatly facilitates subsequent computer manipulation of the representations. A significant part of the structural manipulations which must take place can be treated algorithmically. Development of such algorithms has reached a highly sophisticated state; these developments represent a strong foundation on which to build subsequent procedures which rely on judgmental knowledge, or rules, to arrive at scientifically meaningful conclusions.

The "knowledge engineering" aspects represent a set of similar problems in system design shared by all three projects. Here the concept of community building and sharing of ideas, factors inherent in SUMEX as a resource, play an essential role in allowing the projects to learn from one another and from AI programs in other major areas.

The biochemistry projects have as a common goal the development of interactive programs which act as problem-solving assistants to an investigator. In order to be useful to a wide community, such programs must be capable of assisting in the solution of a variety of real scientific problems. Here SUMEX is indispensable. The resource provides many facilities for access to programs, for recording of terminal sessions, for rapid exchange of messages about problems and their solutions, and for development and export of versions of programs for use in other laboratories.

Using the DENDRAL project as a concrete example, SUMEX has been used for program development and application to many structural problems of the DENDRAL group and their collaborators throughout the country. Export of the CONGEN program began about eight months ago and already eighteen copies of the program have been distributed to other laboratories. SUMEX will continue to be used for development and for exposure of several new programs (adjuncts to or successors of CONGEN) to structural problems here at Stanford, with export taking place after developing confidence in the programs. In addition, new research projects have been undertaken with a small number of collaborators. These persons are interested in development of new techniques for structural analysis, especially in the area of stereochemistry. Network access to SUMEX has been provided so that development of the techniques themselves will take place at one central facility, with the message system providing the primary means of communication between DENDRAL project members and their collaborators. Specific structural problems, for example the conformational studies of Dr. Cowburn at Rockefeller University, come from the collaborators and exemplify the type of problem which the programs must be capable of solving in order to be useful to the community of persons engaged in related research.