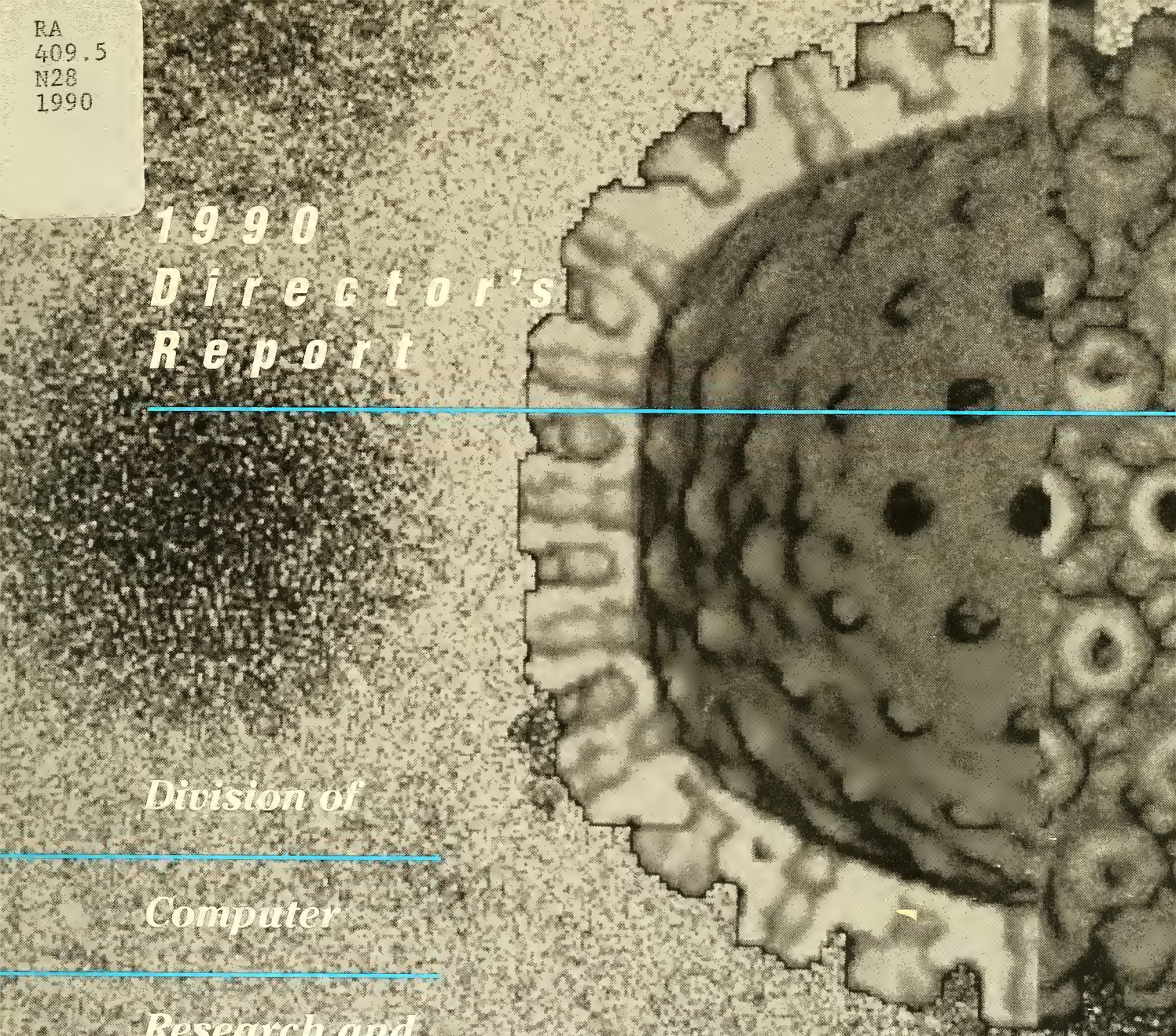
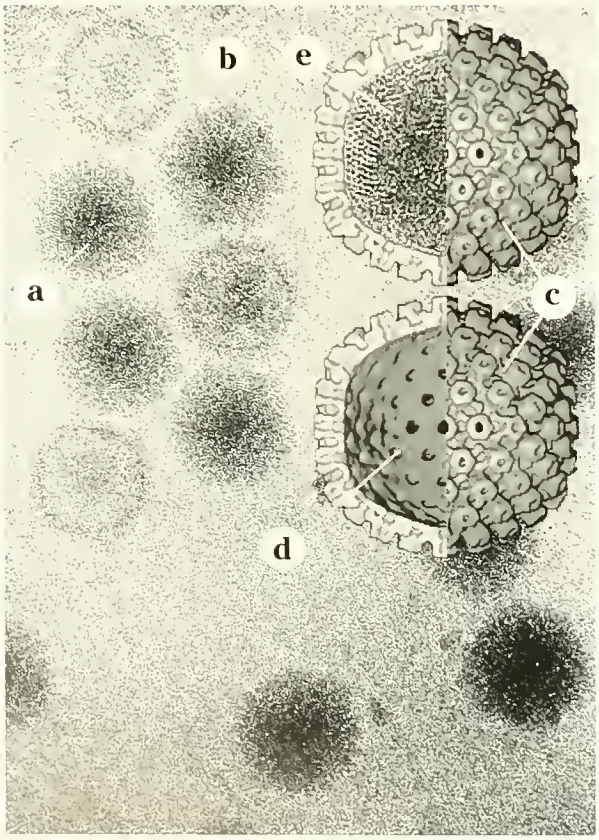


1990 Director's Report

Division of Computer Research and Technology

U.S. Department of Health and Human Services

Public Health Service
National Institutes of Health
Bethesda, Maryland 20892



Cover *(full view)*

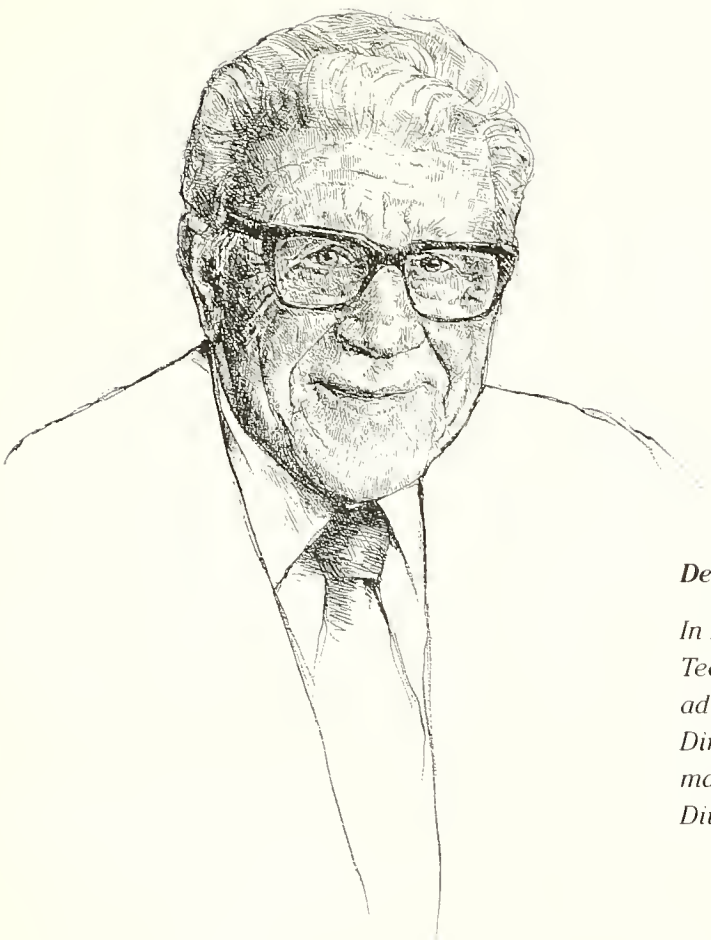
Background: cryo-electron micrograph of herpes simplex virus (type 1) nucleocapsids. (a) DNA-containing (full) capsid, and (b) DNA-free (empty) capsid. Foreground: three-dimensional computer reconstructions combining information content from 30 capsid images. (c) outer capsid surface, (d) inner surface of capsid shell, (e) encapsidated DNA. (Image processing by DCRT and the National Institute of Arthritis and Musculoskeletal and Skin Diseases in collaboration with the University of Virginia, Charlottesville, and Purdue University.)

1990 Director's Report

Division of Computer Research and Technology

U.S. Department of Health and Human Services

Public Health Service
National Institutes of Health
Bethesda, Maryland 20892



Dedication

In 1990, the Division of Computer Research and Technology bid farewell to its retiring leader, advocate and friend, Dr. Arnold W. Pratt. This Director's Report is dedicated to "Scotty" for his many and outstanding contributions to the Division over almost three decades.

The Division of Computer Research and Technology (DCRT) has primary responsibility for incorporating the power of modern computers into the biomedical programs and administrative procedures of NIH. DCRT serves as a scientific and technological resource for other parts of the Public Health Service, and for other Federal organizations with biomedical and statistical computing needs.

DCRT programs focus on three primary activities: conducting research, developing computer systems, and providing computer facilities.

The ***Office of the Director*** *provides overall program direction for DCRT, and serves as a central NIH focus for automated data processing policy. In addition, the Office sponsors research and development work in computational bioscience.*

The ***Physical Sciences Laboratory*** *conducts research in mathematical theory and practical instrumentation to explain biological phenomena in terms of chemistry and physics at the subcellular molecular levels.*

The ***Computer Systems Laboratory*** *provides consultation and collaboration in the design and implementation of specialized computer systems for laboratory and clinical applications.*

The ***Laboratory of Applied Studies*** *relates mathematics, statistics, and computer sciences to such biomedical problems as ECG analysis, evaluation of physiological systems in health and disease, and estimation problems in laboratory medicine.*

The ***Laboratory of Statistical and Mathematical Methodology*** *provides statistical and mathematical help in the computer analysis of biomedical data and offers statistical and mathematical packages for users.*

The ***Computer Center Branch*** *designs, implements, and operates the NIH Computer Center and provides assistance, training, and technical communications to nearly 17,000 users.*

The ***Personal Computing Branch*** *provides guidance and support to scientists and administrators throughout NIH in the effective use of personal workstations, local area networks and associated automation technology. PCB works closely with other DCRT labs and branches to monitor these rapidly changing technologies with a view toward the technical requirements of the user community so that these needs are accommodated in the development and support provided by DCRT.*

The ***Data Management Branch*** *serves as the central systems analysis, design and programming resource for data processing projects relating to scientific, technical, management, and administrative data.*

From the Acting Director

Computing at NIH has been a vital and vigorous activity through the years of progress since 1966 when Dr. Arnold W. Pratt became the Director of DCRT. This annual report stands as clear testimony to the work of all members of the Division in moving the enterprise forward in another year.

DCRT staff met the demanding daily needs of thousands of NIH scientists and administrators. They have confronted a continuing challenge of change in computing technology and in biomedical research. And they will move ahead in the coming decade under the new Director expected to be named early in FY91.

A few brief highlights from the many activities described in these pages reflect the way in which DCRT laboratories and branches keep NIH at the leading edge of computing. This forefront includes the technologies that comprise high-performance computing and computer networking. In the early 1980's, high-performance computing was large supercomputers. Today, it spans a range from large to rather small machines. In the same way, networking technologies have proliferated, and the task for DCRT has been to link those that work well so that each part of NIH can benefit from a piece that fits its needs.

- ***The iPSC-860 Scalable Highly Parallel Computer*** *was installed in September 1990. Its arrival resulted from some three years of careful study of parallel processing options. The effort was led by a senior member of the DCRT Computer Systems Laboratory (CSL) who has a thorough knowledge of computer architectures and many years of hands-on experience implementing innovative systems for biomedical research laboratories throughout NIH. Future work will take advantage of collaboration with the Touchstone Project, which includes other leading centers of computing across the country.*

  The iPSC-860 complements other modern NIH high-performance computers, such as the supercomputer in the National Cancer Institute and the vector facilities and the "minisupercomputer" in the DCRT/NIH Computer Center. Each has its special strengths. Together, they provide a powerful resource for NIH scientists to use as appropriate for their specific research needs. What makes this alliance possible is the evolution of rapid, robust communications among NIH computers of all sizes.

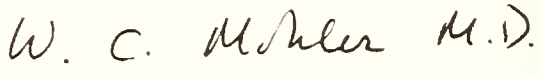
- ***NUnet and RESnet*** *are the latest stages in the development of state-of-the-art networks to link computer facilities and local area networks (LANs) throughout the NIH. In RESnet, the Network Task Group continues DCRT's work on advanced engineering for special NIH networking needs. In NUnet, the Computer Center Branch (CCB) rapidly deploys and enhances an NIH-wide network using available T1 technology.*

NUnet follows more than two decades of CCB experience in providing an increasing variety of services to thousands of users, first via dialup and leased telephone lines and later by external BITNET and InterNet links. In recent years CCB linked the NIH Computer Center facilities to internal DCRT LANs developed by CSL and Personal Computing Branch (PCB) staff and used by all Division laboratories and branches. This undertaking provided a thorough experience with complex multi-network facilities before NUnet was extended across the NIH campus and to outlying NIH buildings.

- ***The Computational Molecular Biology unit** was created as a small group with balanced expertise in molecular biology as well as computing. It took many months to recruit people to strike the balance between a "research" orientation that will keep it scientifically and technically current and a "service" orientation that will help make computational molecular biology useful to scientists throughout NIH.*

This group fills a niche that complements the predominant research focus of DCRT groups like the Molecular Graphics and Simulation unit and the Physical Sciences Laboratory, and the predominant systems focus of groups such as PCB's Scientific Technology Section and CCB's Laboratory Systems Unit.

Emergence of new projects and work groups and change in old ones are inevitable, particularly in fields where both the science and the technology are constantly evolving. DCRT has in the last two decades fostered such evolution through the work of its scientists, engineers, mathematicians, and applications programmers. In addition to many systems for and collaborations with specific scientists and administrators, DCRT staff has provided powerful statistical packages for all NIH users, developed outstanding facilities for molecular graphics and image processing, and helped others to use them. NIH can look for DCRT to continue its innovative approach to biomedical computing into the decade of the 90's.



William C. Mohler, M.D.
Acting Director, DCRT

Contents

OD

Office of the Director

William C. Mohler, M.D., Acting Director

The Office of the Director provides overall program and management direction for DCRT. The Director, Associate Director, Assistant Director and Executive Officer work together as the immediate Office of the Director.

Two other offices supplement the work of the DCRT laboratories and branches:

The **Office of Administrative Management** (OAM) provides administrative and managerial support for the work of DCRT. OAM includes the Administrative, Personnel, Financial Management and Project Control Offices.

The **Office of Scientific and Technical Communication** (OSTC) serves as the central source of information about DCRT activities and about computing and related disciplines. It includes the DCRT Library and the DCRT Information Office.

Also within the Office of the Director is the **Equal Employment Opportunity Office** (EEO), which manages a full EEO program for the Division. The office serves as the focal point and advisory for all activities relating to the equal employment opportunities of DCRT employees and applicants. The EEO Officer maintains a close working relationship with the NIH Division of Equal Opportunity and other components concerned with minority and women's issues.

The Division's role in NIH Information Resources Management (IRM) activities is shared among members of the immediate Office of the Director, but responsibility for NIH ADP planning, budgeting, and security still resides in the Policy Coordination section.

The Office of the Director also has a computational bioscience component that applies computing technology to biomedical research to solve problems in macromolecular structure representation and modeling for collaborating NIH scientists and visitors.

Network Task Group

David C. Songco, Chief

Networking is critical to DCRT's role of supporting computing at NIH. Several groups within DCRT are actively working to develop and implement network technologies for the NIH community. The Network Task Group (NTG) was established in May 1989 to bring increased emphasis on DCRT networking activities in three areas:

- design and implementation of a backbone network to link local area networks (LANs) on campus
- guidance and support for locally managed networks
- design and support for DCRT networks.

The NTG currently consists of six electronic engineers who specialize in the design, development, and implementation of networks using state-of-the-art communications technology.

Background

The establishment of high-speed communications between computers is recognized as an essential resource for scientists and administrators at NIH. DCRT has been designing and implementing a campus-wide backbone network for several years to meet this need.

In January 1990 the NTG presented a report to the NIH Network Policy Board recommending that the original design of a broadband campus area network project be replaced by a new two-pronged DCRT approach to connectivity at NIH.

The New Plan

The new plan called for assigning the task of providing primary connectivity among all buildings on campus to the DCRT Computer Center Branch (CCB) and refocusing the development efforts of the NTG on the task of providing high-speed connectivity to selected sites across NIH in support of biomedical research.

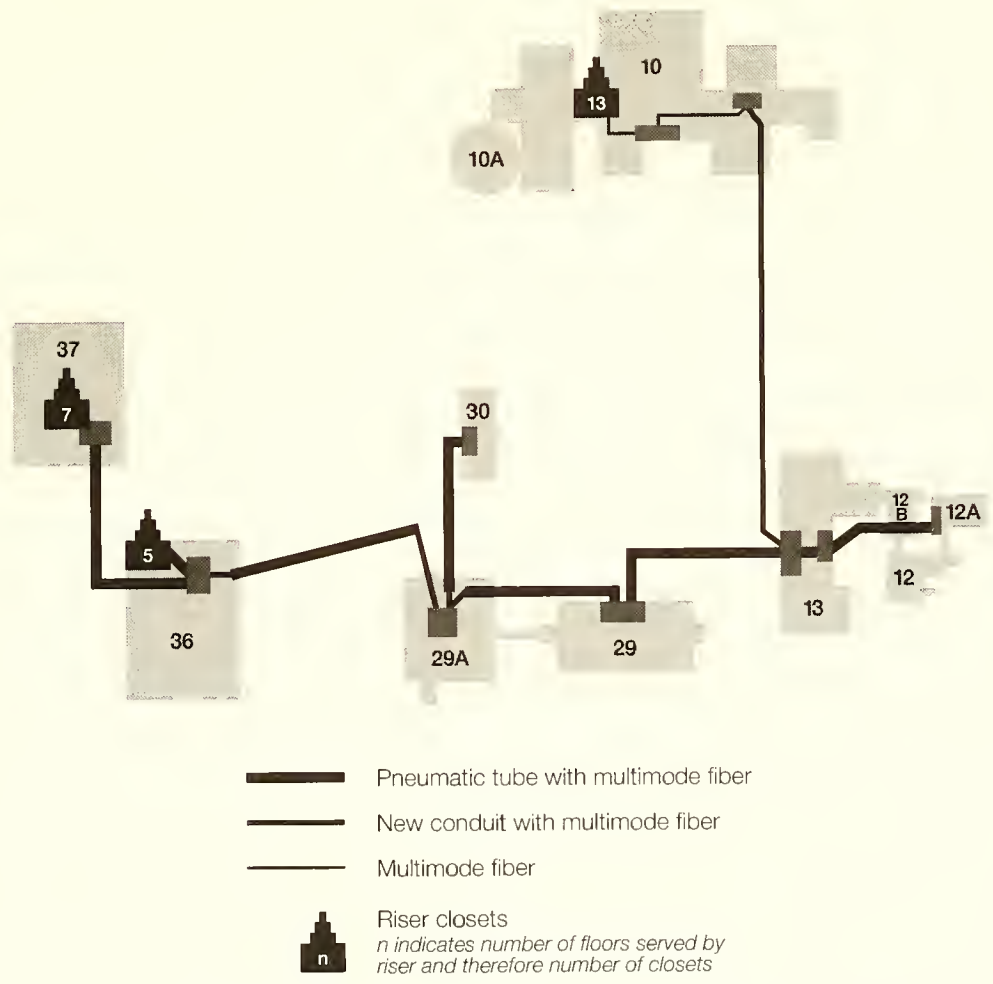
The Computer Center Branch responded to the difficult challenge of linking all local area networks on campus by expanding the scope of the "T1" telecommunications network already under development to provide connectivity to the DCRT mainframe computer facility. The initial phase of this new campus-wide backbone, known as "NUnet" (NIH Computer Utility Integrated LAN and Workstation Support Network) operates at 1.5 million bits per second and is projected by CCB to be completed by December 1990. FY90 costs of NUnet were covered by existing funds allocated for the original backbone project. Additional information regarding NUnet can be found in the Computer Center Branch section of this report.

The NTG then concentrated their engineering talent on developing high-speed connectivity in support of biomedical research. In February of this year, the NTG implemented the first phase of RESnet (Research oriented Network), a network service designed to complement NUnet by providing a high-speed link among selected scientific research sites across NIH. RESnet is linked to NUnet, allowing users to fully participate in campus-wide connectivity, and is available for those applications requiring higher speeds.

Research Oriented Network (RESnet)

Currently RESnet operates as a 10-million bits-per-second fiber optic Ethernet backbone. The backbone spans eight buildings on the NIH campus: 12A, 10, 13, 29, 29A, 30, 36 and 37

Figure 1. RESnet: Ethernet over Fiber Connections April 1990



(Figure 1). The RESnet architecture is designed as a counter-rotating dual ring, similar to the 100 million bits-per-second Fiber Distributed Data Interface (FDDI) standard. If the primary ring should fail due to a broken cable or optical signal failure, traffic is automatically routed to the second ring, adding an extra level of reliability. Because of this reliable design, we have experienced 100% availability of the physical fiber ring since RESnet became operational in February 1990. Routers are connected to the fiber optic backbone transceivers to provide the

interface to local area networks. The 3Com XNS, DECnet and TCP/IP protocols are supported by both NUnet and RESnet routers.

The first major application for RESnet was in linking Silicon Graphics workstations across campus. A total of 27 Silicon Graphics machines were connected via RESnet in FY90. These powerful graphics machines equipped with sophisticated software are well suited to molecular modeling. With the demand for graphics modeling growing and the price of graphics workstations decreasing, it became feasible to place these important tools directly in the hands of NIH scientist and at the same time provide high-speed connectivity among the workstations.

Improvements to RESnet are Planned in FY91

The DCRT Network Task Group is continuing its research and development of higher speed networking technologies. Both the 100-million-bit FDDI (fiber optic) technology and the 10-million-bit Ethernet technology are currently being tested by the NTG. We have installed FDDI equipment and will evaluate this promising technology to determine its suitability as the basis for a standardized high-speed connectivity network.

The architecture of RESnet was designed to permit easy migration to faster technologies such as FDDI. In FY91 parts of the existing 10-million bits-per-second Ethernet will be upgraded to the 100-million bits-per-second FDDI. Migration of the RESnet backbone to FDDI technology will also facilitate the use of FDDI technology at the local level for high-bandwidth applications since these LANS will be able to communicate with each other at full FDDI speeds via RESnet. In FY91 we plan to add six more buildings to RESnet and begin the migration process to FDDI.

Clinical Center Network (CCnet)

The DCRT Network Task Group has been working closely with Clinical Center staff since November 1989 to design and install a backbone network within Building 10. This network, referred to as CCnet, presently exists as an interim fiber optic infrastructure consisting of 28-strand Fiber Optic Cable installed to each of the 13 floors in the west wing. A total of 12 local area networks located throughout Building 10 were connected via CCnet this year. CCnet in turn connects to RESnet in its Network Operations Center located on the B1 level of the building. This room also contains the routers, racks, fiber patch panels, and network management equipment needed to manage and operate CCnet.

CCnet is critical to the success of our DCRT networking plans since nearly 40% of the campus population reside in Building 10. In FY91 DCRT plans to continue its collaboration with the Clinical Center staff and begin the installation of a comprehensive fiber infrastructure throughout Building 10. We estimate that the construction of the LAN riser closets and the installation of the fiber optic cable can be completed within two years.

Support and Guidance for Locally Managed LANs

In addition to their development work on RESnet and CCnet, the NTG engineering staff provide guidance in the design and implementation of local area networks for network planners

within the ICD's. NTG staff have developed and distributed a comprehensive guide to building wiring that details the design of network closets and covers the use of a variety of media including both unshielded twisted pair and fiber optic cable.

This year the NTG continued the test and evaluation of the use of Ethernet over twisted pair and began recommending this strategy for new installations. Unshielded twisted pair cable is gaining popularity as a LAN media because it is the same media used for telephone installations. The NTG staff have developed a plan for rewiring the DCRT buildings 12, 12A, and 12B with a combination of unshielded twisted pair and fiber optic cable. This new cable design will upgrade the DCRT cable plant to the design level that we recommend for new LANs and provide ample capacity for growth.

In FY91 we expect the demand for guidance in the design and implementation of local area networks to increase, especially from the scientific community. This will be driven by the desire to move large amounts of data between computing levels at high speed in order to realize the full potential of the workstation, the server, and the mainframe. To prepare for this challenge, the NTG engineering staff are continuing their skill development by attending local universities and by actively participating in networking conferences and user groups.

The DCRT Local Area Network

The DCRT LAN began as a development project in the DCRT Computer Systems Laboratory in the early 80's with less than a dozen nodes. Since then it has grown to over 250 nodes and is an integral part of the daily operation of DCRT. Support for the DCRT local area network is provided by the NTG in collaboration with the Personal Computing Branch and the Computer Center Branch. In addition, other DCRT organizations assist in this process via their server administrators.

The NTG responsibilities include coordinating the planning and installation of new nodes on the DCRT LAN as well as monitoring the performance of the network. Maintaining the integrity of networks is a complex task that requires the full cooperation of several DCRT laboratories and branches. Network management protocols and the hardware and software tools needed to implement them are still evolving. In FY91 the NTG staff will continue the test and evaluation of products to assist them in maintaining the DCRT LAN and RESnet.

Future plans for the NTG

This past year DCRT has made major progress in networking with the success of NUnet, RESnet, and CCnet. In FY91 the program activities of the NTG will be reviewed with the new DCRT Director to assess the best use of the engineering expertise of the NTG combined with the considerable resources and experience of the Computer Center Branch. Of major importance is the need to develop a coordinated plan for providing the best mix of networking technologies for the NIH and the management, operation, and support of NUnet, RESnet, and CCnet.

Computational Bioscience

Rapid theoretical and technological advances have taken place in the field of molecular biology in the last few years, and rapidly expanding data bases associated with this disci-

pline have dramatically increased the opportunities to apply computational techniques. As a result, the determination and study of nucleic acid and protein sequences and structures have become critically important to the NIH Intramural Research Program, and DCRT has responded with an initiative in computational molecular biology to synthesize engineering, mathematical, and biological expertise into a coherent program of leadership, support and guidance for the intramural program. The program involves implementing an extensive training program highlighting computational tools available to the biologist; acquiring and maintaining several major data bases; providing and extending commercial and academic algorithms, and advising on software and hardware available to the NIH scientists.

The Computational Molecular Biology unit offers consultation, support, and guidance in computational biology to any interested NIH scientist. The staff currently supports the Genetics Computer Group sequence analysis software, as implemented under UNIX on the DCRT Convex C230, as well as a new fast searching algorithm (BLAST) developed by scientists at the National Center for Biotechnology Information. Negotiations are underway to acquire a site license for software with similar capabilities for the Macintosh, and software for other popular workstations is being investigated. A specialized data base, designed to assist researchers involved in examining the "zinc finger" proteins, is currently under construction. The staff is also defining an appropriate research program to maintain currency in the field and provide an environment that will continue to attract good minds to this endeavor. Several promising candidates with similar research interests have been identified, and an effort to identify others is under way.

The Molecular Graphics and Simulation unit (MGS) studied problems of biological significance using theoretical techniques of molecular dynamics, molecular mechanics, modeling, *ab initio* analysis of small molecule structure, and molecular graphics. MGS scientists have focused on the following areas:

- Basic research to provide better understanding of biochemical systems. Emphasis has been on macromolecular simulation for analyzing the structure/function relationship and other properties, computational drug design, and ab initio calculations for determining reaction mechanisms and designing transition state analogs.
- Applied simulation research, usually pursued in collaboration with NIH IRP or extramural experimental scientists. Such research has used molecular dynamics simulations to predict function or structures of peptides and proteins, with application to specific research into the development of vaccines and antiviral agents.
- Development and characterization of new theoretical techniques, often coupled with software and hardware development, involving the generation of new simulation techniques and the systematic testing and evaluation of methods.
- Software and hardware development required to obtain optimal performance for simulation research and to develop new tools for future research. Microcoding commercial processors and engineering-specialized high-performance computers for specific computational chemistry needs were practical manifestations.

- Direct services included support and guidance for using computational techniques; a computing resource for short-term graphics and modeling needs; training and seminars on modeling and simulation methods; and new technology examination and evaluation.

In FY90, 40 Silicon Graphics workstations equipped with molecular modeling software were purchased on a cost sharing basis with scientists throughout NIH: units were placed in laboratories on the Bethesda campus; at the NCI Advanced Scientific Computing Laboratory in Frederick, Maryland; at the NIEHS campus in North Carolina; and at the NIAID Rocky Mountain Laboratories in Montana. The project followed 15 years of molecular graphics system development in DCRT and resulted from an increasing interest in computational molecular structure modeling. As the demand for computational modeling has grown, extending to research in a variety of proteins, protein fragments, and nucleic acids of the viruses associated with AIDS, the dropping prices of graphics workstations has dropped, making large-scale deployment of these tools feasible. Nearly all of the workstations were interconnected to each other and to the Internet, using the high-speed, fiber-optic RESnet. Ten two-day training sessions were provided for some 200 NIH scientists during the year, enabling many to begin useful computational modeling in their laboratories.

During the fiscal year, a plan of collaboration with the Japanese Institute for a New Generation of Computer Technology (ICOT) was established. Known as the Fifth Generation Computer Project, the collaboration involves highly parallel hardware and implementation of a new language with many similarities to PROLOG. The machines may offer novel approaches to several challenging problems. The first application of the collaboration is a relational retrieval system for genetic information.

Research Projects

The Temperature Dependence of the Dynamics of Hydrated Myoglobin: Comparisons of Force-Field Calculations with Neutron Scattering Data

Richard J. Loncharich, Ph.D.; Bernard R. Brooks, Ph.D.

Molecular dynamics simulations of 150 psec in length were carried out on the carboxy-myoglobin protein at 11 temperatures between 20 and 340 K. The simulations attempted to mimic neutron scattering experiments very closely by including a partial hydration shell around the protein. Theoretically determined elastic, quasielastic, and inelastic neutron scattering data were directly compared with experimental values. The inelastic and quasielastic spectra showed that the inelastic peak was shifted to lower frequency than the experimental value, while quasielastic behavior was in good agreement with the experiment, suggesting that the theoretical model is too flexible in the harmonic limit (low temperature) but accurately reproduces high temperature behavior. The average atomic fluctuations showed the protein behaves harmonically at low temperatures. At approximately 210 K, a glass-like transition in atomic fluctuations was seen. Above the transition temperature, the atomic fluctuations exhibited both harmonic and anharmonic behavior. Comparison of protein mobility behavior with experiment indicated the fluctuations derived from simulations are larger in the harmonic region; however, the anharmonic region agrees very well with experiment. Heavy atom dihedral transitions were monitored as a function of

temperature, and trends in the type of dihedral transitions that occur with temperature were clearly visible. Dihedral transitions involving backbone atoms occurred only above the glass transition temperature. Anharmonic motion above the glass transition temperature was in part due to, and possibly dominated by, torsional jumping between substates, and was in accord with recent suggestions from neutron scattering experiments.

The Phase Dependence of Dynamics of Myoglobin: Comparisons of Force-Field Calculations with Neutron Scattering Data

Richard J. Loncharich, Ph.D.; Bernard R. Brooks, Ph.D.

Molecular dynamics simulations were used to probe the atomic motions of the carboxy-myoglobin protein. Simulations of 150 psec in length were carried out on the protein at 100, and 300 K. Vacuum simulations were compared directly to simulations of the partially hydrated protein. Simulations of a cluster of three myoglobin molecules were also carried out, in order to mimic the powder state of proteins. Theoretically determined elastic, quasielastic, and inelastic neutron scattering data were derived from the simulation trajectories and compared. At this point, the results indicate that all vacuum models are more flexible than the hydrated models, with the exception of a vacuum model which used a switching function to truncate nonbonded interactions. All vacuum models showed twice the conformational flexibility of the hydrated model. Future simulations will explore the effects of solvation, partial solvation, the crystalline environment, and other cluster models.

***Ab Initio* Calculations of Carbocations of Substituted Benz[a]anthracene 5,6-oxides**

Richard J. Loncharich, Ph.D.; Bernard R. Brooks, Ph.D.
with Donald M. Jerina (Chief, NIDDK/LBC)

Ab initio calculations were performed on benz[a]anthracene 5,6-oxide, 12-methyl benz[a]anthracene 5,6-oxide, and 1-methyl benz[a]anthracene 5,6-oxide at the STO-3G basis level. Calculations were also carried out using the semiempirical AM1 method. A Boltzmann distribution of the energies of the calculated isomers were in qualitative accord with the trends for methanolysis reactions. The structure of the parent BA 5,6-oxide was nearly planar, while the substituted isomers were nonplanar. The C-6 carbocation was found to be favored over the C-5 carbocation. This was in part due to the 9 kcal/mol resonance stabilization energy of the naphthylic cation over the benzylic cation.

Semiempirical Calculations of Alcohols, Purines, and Substituted DNA Base Pairs

Richard J. Loncharich, Ph.D.
with Steven Musser, Ph.D. (CC)

Semiempirical calculations were performed on alcohols, purines, and substituted DNA base pairs. The calculations used both the MNDO and AM1 methods. The lowest unoccupied molecular orbital energies correlated well with fragmentation patterns in the mass spectrometer.

Cooperativity of Cystine Bridge Formation in Peptides and Proteins

Richard J. Loncharich, Ph.D.; Indira Chandrasekhar, Ph.D.; Bernard R. Brooks, Ph.D.

Free energy simulation techniques have been used to investigate the cooperativity of cystine

bridge formation in Bovine Pancreatic Trypsin Inhibitor, which has been well characterized experimentally. This project, which examines the atomic basis for the cooperativity, involves the use of new software tools and uses a dicystine tripeptide to examine the solvation effects of bridge formation for both cis and trans tripeptides.

Modelling and Simulation of Lipid Crystals and Lipid Bilayers

Indira Chandrasekhar, Ph.D.; Bernard R. Brooks, Ph.D.; Richard J. Loncharich, Ph.D.
with Richard W. Pastor (FDA/CBER)

This project seeks to understand membrane structure, organization and dynamics. Since the study of motion in lipids is critical to an understanding of cell organization and function, the initial aim is to provide a rationale for lipid motion and then to simulate accurately membrane-bound peptides and proteins in the lipid environment.

Current work involves evaluating different energy parameters with regard to their efficacy in simulating lipid crystals (DMPC, DLPE). The initial focus is on varying the non-bonded parameters, namely the charges and the Lennard-Jones constants, since the internal parameters such as bond length and bond angle tend to be somewhat invariant. The analysis of the theoretical crystal involves calculation of the pressure, energy, atomic fluctuations, and deviations from crystal observables, both gross (volume, chain tilt etc.) and detailed atomic (conformation, torsional fluctuations etc.).

Simulations have been set up to test up to 10 parameter sets, both explicit and extended atom. In the former the hydrogen atoms are explicitly included, while in the latter, they are approximated as part of the heavy atom to which they are attached. Future work after optimizing the energy parameters will involve modeling and simulating a lipid bilayer in the fluid or liquid crystalline phase.

The Time Relaxation Behavior Interleukin 1beta: a Simulation Study

Indira Chandrasekhar, Ph D.; Bernard R. Brooks, Ph.D.
with Marius G. Clore (NIDDK/LCP)

Interleukin 1beta is produced in practically all nucleated cells in response to injury or trauma and is known to be a biological mediator in the immune and inflammatory response of almost every organ system. This study examines the time relaxation behavior of the backbone hydrogens of the protein. Nuclear magnetic resonance studies, by Dr. Clore and coworkers of the NIDDK, indicate that certain residues in the protein have extremely slow relaxation times, with two distinct relaxation rates. The crystal structure of interleukin 1beta has been solved independently by three groups and the protein is found to consist of 12 beta strands. This structural information is used in the simulation, which enables the relaxation rates of every residue in the protein to be characterized using the second-order Legendre polynomial and the correlation function to be fit to an exponential, thus providing a rationale for the dynamic behavior of the system.

Harmonic Analysis of Native and Mutant Lysozyme

Indira Chandrasekhar, Ph. D.; Bernard R. Brooks, Ph.D.
with Brian Matthews, Rick Faber (Institute of Molecular Biology, University of Oregon, Eugene)

This study examines the use of harmonic simulation techniques, the hinge bending in

lysozyme. Interest in the problem was kindled specifically by recent crystal data on a mutant lysozyme. The mutant is found to crystallize in two different space groups, in one of which the four distinct conformers with different hinge bends are observed. Harmonic motion (normal mode) calculations are being carried out on the molecule and the energetics of the different residues are being examined to identify "points of stress." The goal is to predict the key residues which control the structure, conformation and stability of a protein.

An extensive study of this kind combined with a systematic examination of the contributions of the different residues is certain to yield significant results regarding stress points in the stability of lysozyme and the mechanics of hinge bending. This work may predict or suggest further mutants that have an interesting behavior. The ultimate goal of this work is to be able to design structural mutants that have desired properties.

Examination of the Role of Cyclic AMP in the Stability and Activity of Catabolite Gene Regulatory Protein

Indira Chandrasekhar, Ph.D.; Bernard R. Brooks, Ph.D.
with Dr. Sankar Adhya, Dr. Susan Garges (NCI/LMB)

The cyclic AMP receptor protein (CRP) or catabolite gene activator protein, as it is also called, is a DNA binding protein that activates or represses the transcription of a variety of operons. The structure and function of CRP has been studied genetically, biochemically and by physical means. The crystal structure of the protein, which exists as a dimer, is known; and it has been deduced from the crystal structure and from fluorescence studies that the binding of cAMP causes a conformational change in the protein. It is further known that in the absence of cAMP, the protein binds non-specifically to DNA, but upon addition of the allosteric affector, becomes sensitive to different proteases.

The conformational change in CRP, induced by cAMP, is being investigated. The protein exists as a dimer in the crystal and has a molecule of cAMP bound to each monomer. This work has minimized the crystal in the presence of none, one and both of the cAMP molecules; in other words, four minimization experiments have been run: (1) with both cAMP molecules, (2) with cAMP attached to monomer A which has a relatively closed conformation, (3) with cAMP attached to monomer B which is relatively open, and (4) with no cAMP molecules present. The results indicate that there is no gross or dramatic change in the structure, such as compactification of the more open monomer B or large movement of the DNA binding helices. The induced conformational change is a local one involving changes in the juxtaposition in the network of hydrogen bonds and interactions in the flexible hinge regions that link the cAMP binding domains to the DNA binding domains. Analyses of these complex interactions are currently in progress.

Simulation Studies of DNA

Indira Chandrasekhar, Ph.D.; Bernard R. Brooks, Ph.D.

Recent measurements by Dr. Parsegian and coworkers of the DCRT of the direct forces of interaction between DNA molecules suggest that the interactions can be modeled in terms of a short-range repulsive force and a long-range electrostatic potential, the later being modulated by configurational fluctuations. A simple

approximate model for DNA is being built to incorporate the measured forces in a simulation that will serve as an interesting test for the energy model and incorporate accurate experimental data. Simulations on solvated DNA are also being carried out at the level of full atomic detail to examine the differences between different DNA parameter sets. Having a reliable DNA parameter set is important for any follow up work for this effort as well as for other simulation efforts involving DNA.

Theoretical Study of Polycyclic Aromatic Hydrocarbon-diol-epoxide-DNA Adducts

Bernard R. Brooks, Ph.D.
with A. Weston (NCI)

In this ongoing study, the binding of diol-epoxides to DNA is being studied by modeling and simulating carcinogenic polycyclic aromatic hydrocarbon-diol-epoxide binding to DNA. This work attempts to answer two main questions: (1) Can the wide difference in carcinogenic activity of similar adducts be explained from a theoretical study? (2) Can the cross-reactivity data between adducts and antibodies raised against different adducts be explained? The initial results of our work indicate that there is a strong correlation between carcinogenic activity and orientation of the hydrocarbon in the minor groove of DNA. This work also provides a possible explanation to observed cross-reactivity data of antibodies raised against one adduct and DNA bound to different adducts: that the antibodies recognize the distortion of the DNA backbone rather than the adduct itself and that structures that cause similar phosphate group movements have high cross-reactivities. The study also suggests that there is a strong DNA sequence dependence for carcinogenic activity of polycyclic aromatic hydrocarbon-diol-epoxides. This work continues with further examination this system with a better level of theory, including a better DNA parameter set, full solvation, and a determination of free energy differences between different adduct conformers.

Harmonic and Langevin Analysis of Large Systems

Dusanka Janezic, Ph.D.; Bernard R. Brooks, Ph.D.
with R. M. Venable (FDA/CBER)

In this ongoing project, harmonic analysis of large systems is being used in a variety of ways to provide better understanding of protein systems. The work has involved performing many different types of harmonic analysis methods to the bovine pancreatic trypsin inhibitor as a trial system. Both harmonic and quasiharmonic analysis calculations have been carried out in several different bases. There has also been extensive analysis of the time behavior of the results from harmonic analysis and a direct comparison with simulation results. Further work involving the Langevin mode analysis of this trial system and other systems is planned in order to: (1) better explain neutron spectral (time of flight) data; (2) better characterize friction and solvent damping; (3) find stress points in macromolecules; and (4) explain (or predict) some site specific mutation data.

Molecular Dynamics Simulations of HIV-1 Protease Monomer in Solution

Frederick W. Carson, Ph.D.; Bernard R. Brooks, Ph.D.
with Richard M. Venable (FDA/CBER)

The intramural effort to develop antiviral agents for AIDS has continued with study of HIV-1 protease. The biological function of this enzyme is to cleave the large precursor gag and pol

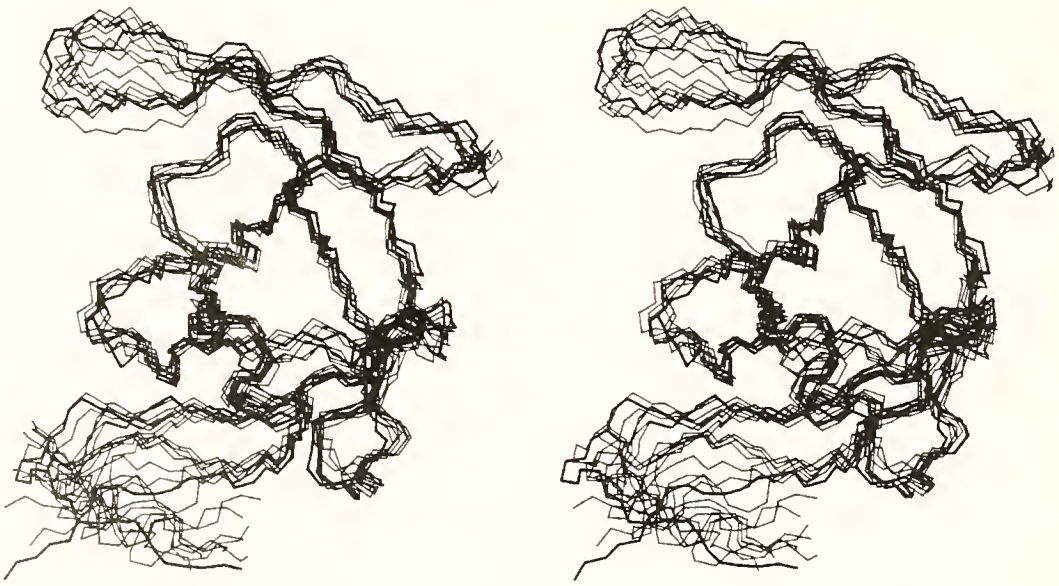
proteins at several sites. Since disruption of this process abolishes infectivity, this enzyme is a suitable drug target. X-ray crystallographic structures are now available as starting points for the analysis of its structure and drug binding. As a first step, molecular dynamics simulations have been used to investigate the reorganization of the structure of a 99-residue monomeric subunit of the HIV-1 protease when dissociated from the native crystallographic homodimer in water. This subunit displayed considerable flexibility in the interfacial region, with some conformational realignment occurring. However, the gross structure did not change much over a period of 100 picoseconds at 300 K (Figure 1). These results imply that the x-ray crystallographic structure of the native dimer may be a suitable starting point for the design of drugs intended to disrupt the dimerization process. Further simulations and detailed analysis of the results are in progress.

Theoretical Analysis of Inhibitor Binding to the Active Site of HIV-1 Protease Using Molecular Dynamics and Free Energy Perturbation Approaches

Frederick W. Carson, Ph.D.; Bernard R. Brooks, Ph.D.
with Richard M. Venable (FDA/CBER)

In a second investigation involving the AIDS HIV-1 protease, work is approaching an examination of the interactions between inhibitors and the protein in a solvated system. Attempts to model the closing of the enzyme's flaps on an inhibitor bound at the active site of the dimer in

Figure 1. Overlay of mainchain atoms from the simulation of the HIV protease monomer



the presence of a single water molecule yielded a variety of structures, some of which resembled the x-ray crystallographic structure. The flaps of HIV-1 protease always seem to close similarly in crystal structures and are also sufficiently flexible to form various hydrogen bonding arrangements which can accommodate a number of different structures of substrates and inhibitors. The similarities between the simulated flap closing and the crystal structure of a complex with an inhibitor suggest that, with further refinement, this approach should prove useful in drug design. In future simulations, the behavior of the flap regions in an aqueous environment will be examined in detail to try to identify their role in specificity. By varying the inhibitor, the key interactions which account for the specificity of this protease may be identified.

In addition, a free energy perturbation study of inhibitor binding has begun. This method is particularly useful when applied to structurally similar inhibitors and can quantitatively assess subtle, but important, interactions with the enzyme. Data from the literature are now available to test the reliability of this method.

Model for the Leucine Zipper of Human Immunodeficiency Virus Type 1 Reverse Transcriptase Using Molecular Dynamics

Frederick W. Carson, Ph.D.
with Richard M. Venable (FDA/CBER); Samuel H. Wilson (NCI/LB); Jean Baillon (NIDDK/LBC)

The fourth AIDS project involves reverse transcriptase, another critical enzyme produced by the AIDS virus. This protein contains an unusual variant of the leucine zipper motif, a dimerization domain found in a number of DNA-binding proteins, such as proto-oncogenes. Leucine zippers are thought to aggregate as helices which form parallel dimers. This possibility is being explored both experimentally and through molecular modeling, including molecular dynamics simulations.

Development of New Theoretical Methods for Studying Macromolecules

Bernard R. Brooks, Ph.D.; Martin J. Field, Ph.D; Richard J. Loncharich, Ph.D.
with R. M. Venable (FDA/CBER)

One of the most interesting and important areas of our research involves the development of new methods. This ongoing work involves the development and characterization of new techniques that have not been tried before. Ongoing efforts in this area include new methods for performing free energy perturbation calculation which allow connectivity to change and have a higher accuracy than existing methods, methods for treating solvent implicitly to provide for hydrophobic affects without the explicit inclusion of many water molecules, methods for properly treating electronic polarization in molecular dynamics simulations, and the development of a more accurate flexible water model. Most of these methods are being developed to provide tools for current needs as well as tools for applied projects in the coming years.

GEMMSTAR, a Two Gigaflop Computer for Macromolecular Simulations

Bernard R. Brooks, Ph.D.; Randall Heck
with William Gandler (DCRT/CSL) and Star Technologies Inc. (under a Cooperative Research and Development Agreement)

Currently in the development phase, this three-year project will result in a processor that should outperform today's supercomputers (Cray X/MP) by a factor of 20 to 30 for a specialized application of macromolecular simulation.

The heart of this computer will be an Application Specific Integrated Circuit (ASIC). Under a collaborative agreement with an industrial sponsor, Star Technologies Inc., DCRT plans to design, develop, and build five prototypes of a direct translation into hardware of the existing software package, Generate, Emulate, and Manipulate Macromolecules (GEMM), developed by DCRT. Design goals are for an extremely favorable cost/performance ratio. Intended for quantity production, the machine will calculate the compute intensive pair interaction potential found in all molecular simulation codes, including a van der Waal term and an electrostatic energy term. Additional terms have been added to make this machine more responsive to future algorithms.

Publications

Brooks, B.R.: Molecular dynamics for problems in structural biology. Nobel Symposium Proceedings, *Chemica Scripta* 29A, 165 1989.

Loncharich, R.J.; Brooks, B.R.: Long range truncation effects in protein molecular dynamics simulations. *Proteins* 6, 32 1989.

Han, K.H.; Syi, J.L.; Brooks, B.R.; Ferretti, J.A.: Solution conformations of the B loop fragments of the human transforming growth factor and epidermal growth factor by ^{1}H nuclear magnetic resonance and restrained molecular dynamics. *Proc. Natl. Acad. Sci.* 87, 2818-2822 1990.

Loncharich, R.J.; Brooks, B.R.: The temperature dependence of dynamics of hydrated myoglobin: comparison of force field calculations with neutron scattering data. *J. Mol Biol.* 215, 439-455 1990.

Weston, A.; Newman, M.; Mann, D.; Brooks, B.R.: Molecular mechanics and antibody binding in the structural analysis of polycyclic aromatic Hydrocarbon-diol-epoxides-DNA adducts. *Carcinogenesis* 11(5), 859-64 1990.

Loncharich, R.J.; Brooks, B.R.: Temperature and phase dependence of protein dynamics: a simulation study of myoglobin. *Proceedings of the IUPAB Satellite Symposium on "Expanding Frontiers in Polypeptide and Protein Structural Research,"* Whistler, British Columbia, July 23-27, 1990.

Field, M.J.: Simulated annealing classical molecular dynamics and the Hartree-Fock method: the NDDO approximation. *Chem. Phys. Letts.* 172, 83 1990.

Brooks, B.R.; Loncharich, R.J.: The development of a high performance computer for macromolecular simulations and the exploration of the temperature dependence of protein dynamics. In: Ikehara, M. (ed.) *Protein Design in Basic Research, Medicine, and Industry.* Springer-Verlag, New York, 9-15 1990.

Patents

Brooks, B.R.: GEMMSTAR: A two-gigaflop computer for molecular simulations. U.S. patent pending number 261,304 filed October 24, 1988. Continuation in Part filed November 1989.

Office of Administrative Management

Marian L. Dawson, Chief

The Office of Administrative Management (OAM) provides administrative support to the DCRT by performing resource allocation, program planning and evaluation, and policy and legislative analysis. The office is organized functionally into sections for administration, finance, and personnel.

During the past year, the Office of the Chief provided staff assistance for major procurements within the Division and prepared several Agency Procurement requests. New electronic software packages were acquired and tested to automate the preparation of solicitations more fully, and forms preparation software packages were tested for efficiency in completing standard government forms.

The Administrative Management Section devoted continuing effort to the division's procurement process. Centralized DELPRO ordering for small purchases was implemented in November, and all small purchases were assigned to one staff member. The administrative staff successfully converted to on-line travel orders and anticipates the implementa-

tion of on-line travel vouchers in the first quarter of FY91.

The Financial Management Section carried out annual budget functions for DCRT, furnishing a variety of monthly reports to program managers to inform them of their monthly budgetary status. The office also participated in the annual rate setting process for Divisional Service and Supply Fund activities and the ITS Budget submission.

The Project Control Office processed requests for new accounts and users of data processing activities, opening 183 new accounts and registering 2,660 new users. This year, the Project Control Office added over 600 users to the new NIH Convex computer, in addition to registering users on the IBM and DEC-system 10. The office completed its annual update of information on over 3,300 accounts and over 19,700 users currently in the Project Accounting System.

The DCRT Personnel Management Section advises and assists management in providing and utilizing human resources to accomplish the goals of the DCRT. The section is responsible for conducting the personnel management program, including staffing-recruitment services, classification and position management, employee benefits, training, employee relations, conduct and ethics.

Fiscal year 1990 was a notable year for the Personnel Office, beginning with a full-scale audit of the entire personnel management program by the Division of Personnel Management. This process consisted of a review of all records, delegations, classification decisions, personnel actions and procedures over the past two years. The audit concluded that this office was in full compliance with all requirements.

DCRT recruitment activity continued at an active pace, aided by a number of special hiring authorities especially for entry-level professionals. Personnel Office staff attended several career fairs and contacted numerous colleges and technical schools in an escalating bid to attract new employees. At the first NIH Job Fair, the personnel office successfully introduced a computerized slide presentation, developed by members of the Personal Computing Branch, which enabled the Personnel Office to showcase current vacancies to prospective employees. The slide presentation was also used during orientation sessions to provide new employees with up-to-date information on the Division.

Once again, DCRT augmented staff by utilizing the summer program, hiring approximately 25 university and high-school students for technical, professional and clerical positions.

The staff provided an increasing variety of assistance and information on employee benefits and services—including the new hiring mechanisms for entry-level professionals and several additional award programs—through seminars, meetings, status reports, memos, and follow-up interviews.

The personnel staff organized the first DCRT-wide briefing on Standards of Conduct and Ethics for all covered employees, in order to address the recent changes in legislation impacting federal employees.

DCRT Library

Ellen Moy Chu, Chief

The DCRT Library provides information resources in computer science, mathematics, and statistics, along with computer applications in biomedical sciences, engineering, information

science, and management. It pursues collaborative testing of new commercially available computer applications for libraries, integrates useful offerings into its own library LAN, and uses a full range of telecommunication-based facilities to communicate with its users and other libraries.

The multimedia collection and information services support other parts of NIH as well as DCRT. This year, over 70 percent of new borrowers were other NIH staff who represented 55 percent of registered borrowers. Library staff processed acquisition or interlibrary borrowing requests of 1,423 books, reports, articles, journal issues, and subscriptions. Much of the increase over last year's total resulted from an increase in DCRT staff interlibrary loan requests.

A new data sheet replaced the Library brochure to update information about services, including electronic mail and network features. For NIH Computer Security Awareness Day, a bibliography was prepared of publications available in the Library. The monthly DCRT LIBRARY NEWS expanded beyond the acquisitions listing to include sections on special journal issues, news about quotes and interviews of DCRT staff in the literature, recent DCRT staff publications, and Library news.

Fiscal year 1990 was a year of numerous staff presentations about information systems recently installed in the Library. The Compact Disk Read Only Memory (CD ROM) technology program for NIH Office Technology Coordinators included hands-on demonstrations. At the March Campus Users Research Exchange (CURE) meeting of NIH 3COM LAN (local area network) users, staff reviewed experiences and technical solutions in networking and internetworking CD ROMs. Beyond NIH, the Library Chief discussed DCRT Library experiences in using LANs to disseminate information services at both the Washington, D.C. FEDLINK Conference on Making Library Automation Choices and the Special Libraries Association Annual Meeting in Pittsburgh, PA.

Another area of major activity was end-user training for self-service searching. Most of this training was conducted on demand, while the user was running an application. Scheduled group seminars included an orientation program describing Library resources and information systems, and a seminar on the Library's most popular CD ROM product, a monthly full-text publication covering PC and Macintosh literature.

The Library's LAN provided DCRT staff network access to the on-line catalog, library information files, and electronic mail. Licensing arrangements allowed network access to three CD ROM publications. DCRT staff, in their own offices, can consult these full-text systems to answer telephone queries from NIH staff or to use in their own project work. Menu systems installed for Library staff and for a public user station simplified LAN navigation and operations.

The Mail Gateway created by the Computer Center Branch provides access to stimulating national library discussion forums on INTERNET. These delivered current information about innovative library developments and lively debates about policy issues raised by new technologies. The forums provided another source for Library staff to seek answers and refer questions about emerging technologies and software systems. BITNET mail inaugurated an international source of reference queries for the Library to answer.

New releases of the on-line catalog and circulation modules extended system functions to allow users to search for publications using Boolean operators to combine search elements. Circulation report capabilities allowed improved monitoring and analysis of usage. Library automation work continued in the barcode labeling of books for the on-line catalog and circulation system. Staff recalled 1,400 items to label and enter into the new system. When completed, this project will close out the old batch circulation system.

Staff development included extensive basic training for a new librarian. The network administrator and backup administrator continued to attend DCRT Server Administrator meetings, CURE meetings, and formal courses. Introduction of self-service searching systems for end users entailed mastering numerous search engines and user interfaces. While staff learned how to search these various products, user aids were produced to simplify and shorten the learning curve for others.

Research Projects

Information Retrieval Systems on Local Area Networks

Ellen Moy Chu

The project is developing local area network (LAN) access to DCRT Library information systems. Network licensing for three CD ROM publications allowed FY90 implementation of multiuser network and internetwork access on the DCRT and Library LANs. Experiences were shared with the NIH Office Technology Coordinators, CURE, federal librarians at the FLICC/FEDLINK Conference on Making Library Automation Choices II, and librarians at the Special Librarians Association 81st Annual Meeting. Consultation services were provided to many NIH staff selecting and installing CD ROM hardware and software. Future plans include investigation of multimedia CD ROMs, Macintosh network access to library information systems, and advisory activities for a proposed NIH public network.

Publications and Presentations

Chu, E.M.: The electronic library: using local area networks for information dissemination. *The Information Professional: An Unparalleled Resource,* Papers Contributed for the 81st Annual Conference of the Special Libraries Association, June 9-14, 1990. Special Libraries Association, pp. 99-105 1990.

Chu, E.M.: Moving toward the electronic library. *FLICC/FEDLINK Conference on Making Library Automation Choices II,* February 22, 1990 Washington, D.C., FLICC/FEDLINK (in press).

Policy Coordination Section

John M. Campbell, Chief

The Policy Coordination Section has responsibility for ADP policy coordination and reporting functions for NIH. These include ADP procurement reviews; the information technology systems budget; ADP annual planning, administrative systems, inventories and hardware; Federal Information Processing standards (FIPS); and ADP systems security. The Section also provides staff support to the Director, DCRT, and the Associate Director for Administration, NIH, for Information Resources Management (IRM) activities.

The information collection process for the annual Information Technology Systems budget has been automated. This allows for budget information from the ICD's to be collected, compiled, and updated more efficiently. NIH

budget information is transmitted to the Public Health Service (PHS) over the network. Transmittal of the source documents reduces the chance of error and allows PHS to receive the information quickly.

Awareness is one of the most important aspects of a computer security program. The Policy Coordination Section held a Computer Security Awareness Day in March 1990 featuring lectures on personal computer viruses to the NIH community. In addition, vendors made presentations regarding their security products. Awareness Day attendees found the lectures and presentations informative and a helpful way to present this information NIH-wide. Another Awareness Day is being planned for FY91. The Section also prepared a videotape about computer security at NIH that was shown at the Awareness Day and has been distributed to ICD Security Coordinators and Executive Officers; in addition, other DHHS components have also requested copies. Other methods that the Policy Coordination Section will be using in FY91 to present computer security information to users include theme posters and desk-to-desk calendars.

An automated Standard Risk Protocol was developed for completing risk assessments and risk analysis required by the Computer Security Act of 1987. This personal computer application is useful for the reviews required in the management of Information Resources (IRM). This automated process is designed to measure the level of security in an application system/facility and to provide management with guidelines for implementing cost-effective safeguards to improve the security risk level. The Standard Risk Protocol consists of several modules. The completed modules are currently in use at NIH, other DHHS components, and a number of other Federal agencies. The Protocol has been accepted as a cost-effective approach to managing ADP security.

In FY90 the summary risk assessment module was completed and is being prepared for distribution. The summary risk assessment module collects and stores data relative to risk analyses, contingency plans, reviews, and certifications. Cross-summarization of the data relative to systems, facilities, and equipment provides management with the capability of correcting vulnerabilities. Also in FY90, updates were made to other parts of the Protocol and the capability to put all of the Protocol on a 3.5-inch floppy was added. Updates and floppies are being prepared for distribution. In the future, the National Technical Information Service, Department of Commerce, will handle distribution of the Protocol to other Government agencies.

DCRT Information Office

The DCRT Information Office continued its involvement in many events and activities affecting the division, the NIH community, and the general public in FY90.

As always, it was a busy year for press involvement with inquiries coming from a number of publications including the *Washington Post*, the *Washington Times*, *Government Computer News*, *Federal Computer Week*, *Science News*, *Omni*, *Computer World*, *Datamation Magazine*, *Personal Computing Magazine*, *Electronic Engineering Times*, *Communication Week*, and *MIS Week*.

In addition, the Information Office responded to more than 1,000 telephone inquiries and

distributed nearly 10,000 publications on DCRT and NIH programs. As part of its public relations effort, the office placed more than 20 stories in the *NIH Record*.

As in previous years, the Information Office devoted a large segment of its resources to producing new editions of *Computing Resources* and the *DCRT Annual Report*, that together present a comprehensive view of the many activities within DCRT. The *Annual Report* was given a new look to reflect the technological advances of the division. The Information Office will continue to improve the quality of materials produced to describe the Division's activities and accomplishments.

DCRT has joined the Clinical Center and the National Library of Medicine in planning the opening of the Computers in Medical Research exhibit providing computer graphics, pictures, and video footage.

The office also worked with the Macintosh group and the Personnel Office to develop an Entrance-On-Duty Macintosh program to be used for employee orientation and recruitment. To keep up with current technology, the office added a computer graphics system, color printer and ScanJet Plus to its facilities available for use by DCRT employees.


PSL

Physical Sciences Laboratory

George H. Weiss, Ph.D., Chief

Members of the Physical Sciences Laboratory (PSL) perform research and offer consultation services in a variety of areas of applied mathematics, biophysics, chemical physics, molecular graphics, and statistical physics. Four permanent professionals and five postdoctoral fellows make up the staff; in addition both visiting scientists and scientists in other NIH institutes collaborate with members of the PSL.

Although much of the research is theoretical, two staff members are engaged in experimental work in their own laboratory facilities: Dr. Ralph Nossal at the University of Maryland, and Dr. Adrian Parsegian in collaboration with members of NIDDK. Dr. Nossal is presently setting up an optics laboratory jointly with members of Biomedical Engineering and Instrumentation Branch (BEIB) to work on a number of applications of modern optics that are related to diagnostic techniques. The different research efforts combine investigations of the physical bases of chemical and biological phenomena with the development of appropriate computational techniques. A typical project under current study includes the development of a theory of photon migration in turbid biological media, with the aim of interpreting data from laser scattering experiments. This project involves members of the PSL and other scientists at the NIH, in addition to a number of visiting investigators in both the experimental and theoretical phases. An important set of results obtained in the last year is the development of a consistent theory for the diffusion of light in a heterogeneous tissue characterized by different absorptivities and diffusion constants. A second project combining both theory and experiment is that of elucidating the nature of forces that determine the structure of biological materials. A particular focus of this set of investigations is on the role of hydration forces in maintaining the separation between different biological structures.

Dr. B.K. Lee is developing a research program for the study of protein folding, seeking to predict protein configurations from a knowledge of the protein sequence, using techniques from statistical mechanics and physical chemistry.

While each study originates in a biological, chemical, or physical context, the techniques used often involve considerable computation, requiring the development of new sophisticated computer algorithms which quite often find a much broader range of application.

The work of PSL members is recognized both nationally and internationally, as evidenced by the large number of invited seminars given by Laboratory members at universities, research institutes, and international meetings. Dr. G. Weiss served as co-chairman of a meeting on Chemistry and Mathematics held at Texas A. and M. University, and is organizing an international meeting to be held next year at NIH on Models for Anomalous Reaction Rates.

Research Projects

Instrumental Analysis

George H. Weiss, Ph.D.
with U. Shmueli, Ph.D. (Tel Aviv University, Israel)

The conversion of crystallographic data to information about the atomic configuration in different molecules requires the solution of probabilistic problems that relate the pattern of scattered X rays to the underlying molecular structure. The techniques used are known in the crystallographic literature either as intensity statistics or direct methods of phase

determination. All of the original theory embodied in algorithms used in x-ray diffractometers relies on approximate solutions to these problems. Computer methods allow at least some of these problems to be solved exactly, which represents one of the projects of the PSL.

One of the major developments in this past year has been a derivation of the characteristic functions needed in the technique of intensity statistics, for analyzing data from molecules crystallizing in 206 out of the 230 possible space groups. The results of this study will be incorporated into programs routinely used for the interpretation of crystallographic data. A second subproject relates to the calculation of the probability density for the so-called "three phase invariant," which is an important crystallographic tool for data interpretation. This project requires a heavy computational effort, and work is proceeding, with a considerable degree of success, on finding approximations to the probability density of the three phase invariant which are not so computationally intensive.

Studies in Applied Mathematics and Statistics

George H. Weiss, Ph.D
with S. Havlin, Ph.D., D. ben-Avraham (Clarkson University); J. Bendler, Ph.D. (General Electric); I. Dayan, M. Gitterman, Ph.D., H. Taitelbaum (Bar-Ilan University, Israel); R. Kopelman (University of Michigan); J. Masoliver (University of Barcelona, Spain)

A large component of this project emphasized the analysis of chemical reaction kinetics, particularly in restricted geometries, where the rates are known to differ from those predicted by classical rate equations. Some effort was also devoted to developing a number of models for transport in a disordered medium.

Continuing earlier work, members of the PSL and a number of collaborators performed extensive work on characterizing the intermolecular distances between reacting components in systems in one and two dimensions. Considerable effort has been devoted to understanding the role of diffusion in reacting systems. A theory has been developed for diffusion in the presence of a molecule with a non-Markovian boundary. This provides a phenomenological model for the study of reactions between molecules with complicated internal states. An international conference on non-classical chemical kinetics has been organized by the PSL and will be held at NIH in March, 1991.

A study motivated by the need to develop more accurate models of photon migration in human tissues has led to the problem of obtaining solutions to the so-called "telegraphers equation" in the presence of trapping boundaries. These solutions have been found for the one dimensional problem. Work is now proceeding to find similar results in three dimensions.

Other topics of investigation relate to the relaxation behavior of a viscoelastic polymer after a stress has been applied, and to the diffusion of gas through a polymer with a distribution of trapping sites. This is also related to the diffusion of gas through the lung.

Interactive Molecular Graphics

Byungkook Lee, Ph.D.
with Dr. Jong Ryul Kim (DCRT/PSL)

GEMM is an interactive molecular graphics package, originally developed by Bernard Brooks, that allows a researcher to examine a molecule by rotation and translation. It continues to enjoy popularity among scientists despite the availability of a number of commercial

packages. The capabilities of the package over the last year have been significantly enhanced as part of the present project. They now include solid surface and ribbon representations and dynamics trajectory play-back. The capability for viewing the trajectory in stereo and the fact that we now have two Silicon Graphics Inc. machines in the laboratory—one for program development and the other for actual use—have been essential for the rapid development of Chorus, our protein folding program (see below).

There have been a large number of requests for copies of the program from outside of NIH. One notable site is the Franklin Institute in Philadelphia. A "tamper-proof" version of the program is running permanently at this site. The program is controlled by a set of buttons that can be punched by children (as well as scientists) and presents fascinating representations of several molecules that can be manipulated in real time by means of dials.

In the past year, a simple mathematical plotting package, GPLOT, was developed for the Silicon Graphics Inc. because of the dissatisfaction with some of the commercial packages available for the Macintosh. Simple to use and easy to modify, GPLOT was developed rapidly by borrowing many of the features from GEMM. The program is being used routinely to produce publication quality graphs using a Postscript laser printer.

Protein Folding

Byungkook Lee, Ph.D.
with Hong S. Kang, Ph.D. (DCRT/PSL)

One of the outstanding problems in biochemistry is whether one can predict the configuration of protein molecules from a knowledge of their sequence. A major project in the PSL involves the incorporation of methods derived statistical mechanics and chemical physics into a protein folding program. The program for this purpose being developed in the PSL, Chorus, has entered the middle stage of its development. The initial stage involved installing the mechanics for making biased Monte Carlo moves in dihedral angle space, which refers to one of the parameters required for the description of protein conformation. This was accomplished toward the end of last year. Preliminary tests run on polyglycine and polyalanine have been quite successful in deriving results in agreement with presently available experimental data. The program folds the latter into alpha-helix and freezes the former into a random glass, in agreement with experimental behavior. One surprise is that the glass transition for the polyglycine occurs at lower temperature and is much sharper than the helix-coil transition for the polyalanine. A quantitative theoretical explanation for this surprising feature does to exist at present.

The present development phase consists of incorporating the side-chain moves. Side-chain packing is a difficult but crucial problem in protein folding because it is here that the sequence-specificity of protein structure arises. Moves for each of the several different classes of side chains are being incorporated one at a time. Incorporation of hydrophobic side chains have been accomplished. The program correctly distinguishes such similar residues as valine, leucine, and isoleucine. Techniques for modifying dihedral angle dynamics are, in the present stage of development, fast but rather crude procedure. Nevertheless they do reproduce experimental results to a considerable

degree. Currently, short polar side chains are being incorporated, to be followed by the long polar groups.

The next stage will consist of incorporation of a more precise treatment of the hydrophobic effect. This will involve more extensive calculations. The final stage will then consist of incorporating procedures that can handle realistically long chains. These include procedures for handling global constraints and decision-making procedures for large block moves.

Biophysical Analysis

Ralph J. Nossal, Ph.D.
with S. Havlin, Ph.D., G.H. Weiss, Ph.D. (DCRT/PSL); R. Bonner, Ph.D., J. Schmitt, Ph.D. (NCRR/BEIP); J. Hofrichter, Ph.D. (NIDDK/LCP); B. Chance, Ph.D. (University of Pennsylvania); S.H. Chen, Ph.D. (MIT); R. Gammon, Ph.D. (University of Maryland); S. Krueger, Ph.D. (National Institutes of Standards and Technology); P. McCormick, M.D. (Henry Ford Hospital, Detroit, Michigan); K. Zaner, M.D., Ph.D. (Boston University Hospital)

Efforts continued toward understanding aspects of photon migration in turbid biological media, with emphasis on analyzing photon propagation through layered biological structures. Results have been applied to various non-invasive medical diagnostic procedures which involve the scattering and absorption of laser light *in situ,* such as the measurement of hemoglobin oxygenation in the brain by time-resolved absorption spectroscopy. Passage of near-infrared radiation through the head was modeled by a composite structure consisting of a layer of optically dense material (skull) overlying a region of lower scattering density but higher absorption (blood-containing tissue). Mathematical expressions were derived for several parameters pertaining to the time course of photon re-emission; and schemes were devised for measuring changes in blood oxygenation, even when organ or tissue geometries seem to preclude precise determination of absolute values.

Essential results of the theory were tested by analyzing published time-resolved data obtained from animal experiments, and by measuring photons propagating within "phantoms" composed of layers of agar gels containing various amounts of dyes and embedded polystyrene latex beads (with Drs. R. Bonner, B. Chance, and P. McCormick). A theory was recently devised to describe photon propagation in convoluted structures, which may have applicability to the optics of highly vascularized tissues (with Drs. S. Havlin and G.H. Weiss).

In related work, methods were developed for detecting movements of scatterers, such as red blood cells or cell organelles hidden within static, optically dense, media. Relationships were established between laser Doppler blood flowmetry and recently developed "diffusing-wave" quasi-elastic light scattering techniques for studying motions of particles in optically dense colloidal suspensions (with S. Havlin). Experiments were performed in which diffusing-wave spectroscopy was used to probe path-length distributions of photons migrating within complex composite structures (with Drs. R. Gammon and J. Schmitt).

An analysis was completed of neutron scattering studies of concentrated solutions of the carbon monoxy (CO) derivative of purified hemoglobin A (with Drs. S. Krueger, J. Hofrichter, and S.H. Chen). Computer simulations indicated that such samples contain limited amounts of aggregated hemoglobin tetramers. The novel techniques worked out in this study are applicable to other investigations of concentrated biological solutions.

Molecular Forces in Cellular Assembly

V. Adrian Parsegian, Ph.D.
with J.J. Zimmerberg, M.D., Ph.D., K. Gawrisch, Ph.D., Sergey Leikin (DCRT/PSL); D.C. Rau, J.J. Kasianowicz, Ph.D., C. Moore, Marcio Colombo (NIDDK); E.A. Evans, Ph.D. (University of British Columbia); D.F. Evans, Ph.D. (University of Minnesota); S.M. Gruner, Ph.D. (Princeton University); R. Podgornik (Lubljana, Yugoslavia, and NIDDK); R.P. Rand (Brock University, Canada); Matthias Winterhalter (Berlin, Germany)

This work is directed toward understanding and integrating the physical properties of forces and energies, necessary for developin a theory of the behavior of matter at the macromolecular or molecular-aggregate level of biological materials. The past year saw a significant extension of the list of substances whose interactions near contact are dominated by powerful exponentially varying "hydration" or solvation forces.

Stiff polysaccharides, particularly xanthan, were found to have a remarkable ability to act as a thickener at very low concentrations and to absorb water under virtually all solution conditions. This material was found to exhibit quantifiable hydration forces in a huge variety of solutions. The correlation with rheological properties, previously measured by others, suggests a way to link molecular features of hydration with the bulk properties that make Xanthan such a useful compound in food processing or oil recovery. Specifically, polysaccharides have been added to the categories of substances exhibiting these forces. Molecules of this relatively neglected class are found in virtually all living matter. Qualitatively new features of their properties in water may have important consequences in understanding phenomena as complex as secretion and cell contact.

Since the earliest development of thermodynamic theory it has been recognized that the work involved in a process, in our case the work of pushing together molecules or membranes, consists of an "entropic" part that includes its temperature sensitivity and an "enthalpic" part for the remainder of the work encountered. We have taken the exponentially varying hydration force and decomposed it into these two parts. A separation of the two components of the work required to squeeze water out from between two molecular surfaces was made quantitative. The entropy described the newly gained freedom of water molecules released due to breaking the weak bonds that hold water near the macromolecular or membrane surface. It appears that each of the components of the interaction energy vary exponentially with distance between molecules. This exponential relationship, now taken as one sign of a hydration force, is of some interest. It is popularly viewed as a showing how the pertubation of water by a membrane or molecular surface involves the successive effect of each "layer" of water disturbing the next one out. The fact that the entropy or order of the water shows this property lends credence to that popular view.

In the study of lipid assemblies, we learned that many phospholipids can change packing from the bilayer form typical of membranes to hexagonal assemblies of the kind thought to be important in secretion and membrane fusion. The energetic factors that control structural transitions—packing of phospholipid hydrocarbon chains, repulsion or attraction of lipid polar groups, electrostatic field energies of lipid charges—have been measured or changed in different ways to see how the transitions can be exploited by cells. For example, small local

changes in pH or ionic strength, under cellular control might be a means for the cell to regulate changes in lipid or even lipid/protein packing. The lamellar or bilayer form of packing, taken to be the backbone of a cell plasma membrane, can then rearrange to push polar ends together and even to merge lipid molecules from previously facing bilayers. By comparing the Nuclear Magnetic Resonance (NMR) and ion transport properties of lipids with and without these dipoles, and by measuring hydration forces of these two kinds of lipids, we have learned that the reorganization of water by membranes is probably due more to structural forces rather than ordering by electric fields.

Further developments require an integrated effort involving theory, computer simulations, data analysis, and experiments. All of these threads are currently being pursued.

Membrane Transport and Fusion

Joshua Zimmerberg, M.D., Ph.D.; V. A. Parsegian, Ph.D., DCRT and NIDDK
with F. Bezanilla, Ph.D., A. Harris, Ph.D. (The Johns Hopkins University)

As a part of our studies to understand how structure and function are related in biological systems, we have tested kinetic models for the behavior exhibited by the protein porin, which is found in the uniquely crystalline outer membrane of *Bordatella pertussis*, the bacterium which causes whooping cough. Isolated by FDA scientists working to produce a better vaccine, porin has been found to form large, conducting channels in planar bilayers which are selective for anions over cations, when incorporated into planar phospholipid bilayers.

In another study, membrane transport was used to measure the dipole potential of lipids. Planar phospholipid membranes were prepared of either DPPC or DHPC, which differ by the presence of the ester-linked carbonyl groups. The investigation demonstrated that the dipole potential cannot cause the hydration force.

Membrane fusion is the pivotal reaction in hormone release, enveloped viral entry, syncitia formation, and fertilization. Despite intense work on lipid, microsomal, and electric pulse fusion, the molecular mechanisms of membrane fusion remain obscure. An investigation was conducted to learn which proteins mediate exocytosis *in vitro*. Essential elements were determined as a strategy towards purification of fusion proteins. A model was created using the isolated planar cortex of the sea urchin egg, where micromolar concentrations of intracellular calcium trigger fusion. The secretory granules were purified and recombined on solid substrate, where calcium-triggered their fusion to form large syncitia. This demonstrated that the molecular machinery responsible for calcium triggered fusion resides on the exocytotic vesicles.

Infection of animal cells by enveloped viruses is accomplished in part when the viral genome is delivered into the cell following the fusion of the cell and viral membranes. This membrane fusion event is catalyzed by a protein; in the case of influenza, the hemagglutinin (HA). In addition to viral-cell membrane fusion, the hemagglutinin also mediates the fusion of cells, provided the molecule exists in sufficient concentration in at least one of the two cell membranes, is activated by appropriate limited proteolysis, and is protonated at pH<5. Cell-cell fusion mediated by viral spike protein was studied using fluorescence dequenching assays on cell populations in a cuvette and on cell-cell pairs using video microscopy. Fusion

between cells expressing influenza virus hemagglutinin (HA protein) and human red blood cells (RBC) was determined by measuring the transfer and dequenching of both a lipid marker and a cytoplasmic marker.

Experiments using cuvettes are statistically advantageous due to a large sampled population of cell-cell interactions. Population studies can be limited by the heterogeneity of the cells with respect to the density of expressed HA protein and binding of RBC's, as well as biological diversity such as the phase of cell life-cycle. To distinguish between the changes in the number of fusing cells and the rate of fusion of each cell, and to determine the range and distribution of fusion kinetics, methods were developed to measure cell-cell fusion on single cells. A flow chamber was invented which allows rapid switching of the buffer solution around cells attached to a microscope coverslip during microscopical observation, making it possible for us to observe fusion events in single cell pairs to be recorded by measuring fluorescence appearing in the unlabeled partner cell.

An examination of fusion at the level of single events added unexpected information.

- Fusion is all or nothing: either a pair will fuse at a given pH or after a pulse, or it will not. We found no evidence of partial fusion.
- Reacidification from neutrality, following failure to fuse, leads to fusion with the same lag time and time-course. Lag time is independent of history of exposure to pH.
- Individual fusion lag times and time-courses are broadly heterogeneous, both in rate and shape of curve.

Analysis of data from the new technique of measuring the exocytotic pore during membrane fusion in the mast cell has continued using improved computational tools that allowed better statistical analysis of the probability histogram of pore conductance, irrespective of time after initiation of fusion. Evidence for steplike increases in pore conductance were found, as if the pore increases its size through the addition of protein monomer to the wall of the pore. The variability in lifetime suggested that the pore is a lipid/protein complex rather than a traditional protein pore.

The studies thus showed that fusion during exocytosis is unique, slower than lipid fusion, with pore widening between 10 and 1000 milliseconds rather than 200 microseconds. Being much more fluid than protein channels, the fusion pore variably visits many different conductances rather than flickering between two fixed conductance states. The fusion pore is now thought to be a lipid/protein complex. Future work will aim at purifying specific proteins and reconstituting biological fusion into planar bilayers, starting with the viral fusion protein, where a small pore or junction seems to mediate fusion.

All phases of our studies include a considerable component of computer use, as well as the development of new algorithms both in the areas of image processing and in the analysis of experimental data. Much new image-processing technology has been developed as a byproduct of this project.

Publications

Agmon, N.; Weiss, G.H.: Theory of non-Markovian reversible dissociation reactions. *J. Chem. Phys.* 91, 6937-6942 1989.

Bonner, R.F.; Nossal, R.: Principles of laser flowmetry. In: Shepherd, A.P.; Oberg, P. (eds.) *Laser-Doppler Blood Flowmetry,* Kluwer Academic Publ. New York, 17-45 1990.

Bonner, R.F.; Nossal, R.; Weiss, G.H.: A random walk theory of time-resolved optical absorption spectroscopy in tissue. In: Chance, B. (ed.) *Photon Migration in Tissues,* Plenum Press, New York, 11-23 1990.

Chandler, D.E.; Whitaker, M.; Zimmerberg, J.: High molecular weight polymers block cortical granule exocytosis in sea urchin eggs at the level of granule matrix disassembly. *J. Cell. Biol.* 109, 1269-1278 1989.

Curran, M.; Cohen, F.S.; Chandler, D.E.; Zimmerberg, J.: An exocytotic pore of variable size mediates fusion in mast cells. *J. Cell. Biol.* (in press).

Gawrisch, K.; Parsegian, V.A.; Rand, R.P.: Membrane hydration. In: Glaser, R.; Gingell D. (eds.) *Biophysics of Cell Surface,* Springer-Verlag, New York (in press).

Gitterman, S.; Shrager, R.J.; Weiss, G.H.: Influence of noise on systems described by a Mathieu equation. *Phys. Lett. A.* 142, 84-88 1990.

Havlin, S.: Multifractals in diffusion and aggregation. Third Bar-Ilan Conference on Frontiers in Physics. *Physica* January 1990 (in press).

Havlin,S.; Larralde, H.; Kopelman, R.; Weiss, G.H.: Statistical properties of the distance between a trapping center and a uniform density of diffusing particles in two dimensions. *Physica A* 169, 337-41 1990.

Havlin, S.; Schoonover, R.; ben-Avraham, D.; Kopelman, R.; Weiss, G.H.: Nearest-neighbor distances in diffusion-controlled reactions modelled by a single mobile trap. *Physica* (to appear).

Havlin, S.; Taitelbaum, H.; Weiss, G.H.: Effects of hard-core interactions on nearest-neighbor distances at a single trap for random-walks and for diffusion processes. *Chem. Phys.* 146, 351-354 1990.

Havlin, S.; Trus, B.; Bunde, A.; Roman, H.E.: On the phase transition in DLA. *Phys. Rev. Lett.* 63, 1189-1190 1989.

Havlin, S.; Weiss, G.H.: A new class of long-tailed pausing time densities for the CTRW. *J. Stat. Phys.* 58, 1267-1273 1990.

Krueger, S.; Chen, S-H.; Hofrichter, J.; Nossal, R.: Small angle neutron scattering studies of HbA in concentrated solutions. *Biophys. J.* 58, 745-757 1990.

Krueger, S.; Lynn, J.W.; Russell, J.T.; Nossal, R.: Small angle neutron scattering method for in situ studies of the dense cores of biological cells and vesicles: application to isolated neurosecretory vesicles. *J. Appl. Cryst.* 22, 546-555 1989.

Marty, A.; Horn, R.; Tan, Y.P.; Zimmerberg, J.: The delay of the Ca mobilization response to muscarinic stimulation. *Proc. Soc. Gen. Physiol. Series* 44 (in press).

Nossal, R.: Quasielastic light scattering from particles diffusing within optically dense media. *J. Phys. Chem.* 93, 6961-6964 1989.

Oppenheim, I.; Shuler, K.E.; Weiss, G.H.: Stochastic processes in physics. *Encyc. Physics* (in press).

Parsegian, V.A.; Rand, R.P.; Fuller, N.L.: Direct osmotic stress measurements of hydration and electrostatic double layer forces between bilayers of double-chained ammonium acetate. *J. Phys. Chem.* (in press).

Podgornik, R.; Parsegian, V.A.: An electrostatic-surface stability interpretation of the "hydrophobic" force reported to occur at 0 to 10 nm intersurface separations. *Chem. Phys.* (in press).

Podgornik, R.; Parsegian, V.A.: Molecular fluctuations in the packing of polymeric liquid crystals. *Macromolecules* 23, 2265-2269 1990.

Rand, R.P.; Fuller, N.L.; Gruner, S.M.; Parsegian, V.A.: Membrane curvature, lipid segregation, and structural transitions for phospholipids under dual solvent stress. *Biochem. J.* 29, 76-87 1990.

Rand, R.P.; Parsegian, V.A.: Hydration forces between phospholipid bilayers. *BBA Biomembranes* 988, 351-376 1989.

Rau, D.C.; Lee, B.; Parsegian, V.A.: Measurement of the repulsive forces between polyelectrolyte molecules in ionic solutions: Hydration forces between parallel DNA double helices. *Proc. Natl. Acad. Sci.* 81, 2621-2625 1989.

Rau, D.C.; Parsegian, V.A.: Direct measurement of the intermolecular forces between counterion-condensed DNA double helices: Evidence for long-range attractive hydration forces. *J. Mol. Biology* (in press).

Rau, D.C.; Parsegian, V.A.: Direct measurement of forces between linear polysaccharides xanthan and schizophyllan. *Science* 249, 1278-1281 1990.

Shmueli, U.; Rabinovich, S.; Weiss, G.H.: Exact random-walk models in crystallographic statistics: Nonsymmetrically bound distributions of structure-factor magnitudes. *Acta Cryst.* A 46, 241-246 1990.

Shmueli, U.; Weiss, G.H.: Exact conditional probability density function of three phases invariant in P1: some theoretical and practical considerations. *Proceedings of the Erice School on Direct Methods* (in press).

Shmueli, U.; Weiss, G.H.: Probabilistic methods in crystal structure analysis. *J. Am. Stat. Soc.* 85, 6-19 1990.

Taitelbaum, H.; Kopelman, R.; Weiss; G.H.; Havlin, S.: Statistical properties of nearest-neighbor distances at an imperfect trap. *Phys. Rev. A* 41, 3116-3120 1990.

Weiss, G.H.: Biography of Elliot Waters Montroll, *Biographical Memoirs of the Natl. Acad. Sci.,* USA (in press).

Weiss, G.H.: Discussion of the paper by Daniel Ben-Avraham, *Chemometrics and Intell. Lab. Syst.* (in press).

Weiss, G.H.: Moviment brownia, difusio, camines aleatoris. *Quaderns* (in press).

Weiss, G.H.; Bendler, J.T.: Random walk model for viscoelastic response of glassy polymers. *Physica A* 168, 592-601 1990.

Weiss, G.H.; Dayan, I.: A generalized radiation boundary condition, *J. Chem. Phys.* 92, 5235-5238 1990.

Weiss, G.H.; Dayan, I.: Kinetic behavior of diffusion systems in the presence of non-Markovian boundaries. *J. Chem. Phys.* 93, 7374-7377 1990.

Weiss, G.H.; Dishon, S.; Bendler, J.T.: Tables of the Laplace transform of the function exp (s^{-B}). *J. Research, NIST* 95, 433-467 1990.

Weiss, G.H.; Gitterman, G.H.: Some comments on approximations to the master equation. *Physica A* (to appear).

Weiss, G.H.; Masoliver, J.; Shuler, K.E.: On the asymmetry of a random walk in the presence of a fiefd. *J. Stat. Phys.*5 58, 643-652 1990.

Weiss, G.H.; Masoliver, J.: Transport equations in chromatography with a finite speed of signal propagation. *Separation Sci. & Tech.* (to appear).

Weiss, G.H.; Schrader, M.E.: Free energy and vapor pressure of sessile drops. 2. Hysteresis. *J. Colloid & Interface Science* 135, 586-589 1990.

Zimmerberg, J.: Fusion of phospholipid vesicles to planar phospholipid membranes. In: Wilschut, J.; Hoeckstra, D. (eds.) *Cellular Membrane* (in press).

Zimmerberg, J.; Marty, A.: Diffusion into patch-clamp recording pipette of a factor necessary for muscarinic current response. *J. Cellular Signaling* (in press).

Zimmerberg, J.; Bezaniffa, F.; Parsegian, V.A.: Solute inaccessible aqueous volume changes during opening of the potassium channel of the squid giant axon. *Biophy. J.* 57, 1049-1064 1990.

CSL

Computer Systems Laboratory

Perry S. Plexico, Chief

The Computer Systems Laboratory (CSL) conducts research and development in computer engineering and computer science to identify, adapt, and apply state-of-the-art computer technologies to NIH biomedical research programs, and collaborates with scientists and clinicians from other Institutes/Centers/Divisions (ICDs) to devise computer-based systems for specific laboratory or clinical research activities involving real-time data acquisition, computation, or control. Technology areas in which the CSL is involved include computer communications, image processing, laboratory automation, laboratory data processing, distributed computing, and software engineering.

The CSL has a staff of 26, plus five others who remain on detail to DCRT's Network Task Group. The staff includes engineers, computer scientists, and computer specialists, as well as people trained in the biomedical disciplines who contribute to CSL's multidisciplinary capability and act as an interface to the NIH scientific community-at-large. The staff exhibits the following characteristics:

- familiarity with the rapidly changing computer hardware and software marketplace to enable selection and purchase of appropriate products and technologies
- a thorough understanding of computer engineering, science, and technology to resolve issues associated with developing computer hardware and software and with integrating individual components into complex systems
- sufficient background in biology, medicine, chemistry, or physics to communicate with collaborating NIH scientists and to develop biomedical computing applications.

During late FY89 and early FY90, CSL reviewed and revised its mission, modifying its internal organization in order to respond better to new and changing computer technologies and apply those technologies more effectively to biomedical science. The CSL carries on its projects through three sections in addition to an office of the chief.

The *Office of the Chief* coordinates planning for new CSL research and development initiatives, coordinates the work of the various sections to ensure cooperation and integration of efforts, and provides overall management of CSL activities.

The *Laboratory and Clinical Systems Section*, under the leadership of Arthur Schultz, continues CSL's traditional role of collaborating with NIH scientific and medical investigators and applying computer technology and engineering to solve specific laboratory and clinical problems. The section typically undertakes automation projects characterized by real-time interfacing between biomedical instruments and computers and system architectures involving workstations and networks.

The *Distributed Systems Section*, headed by Keith Gorlen, investigates the potential of distributed computing technologies, such as distributed file systems, network-transparent graphics systems, and distributed data-base systems to aid biomedical research at the NIH. Its principal focus at present deals with networked, interoperable, open-architecture, distributed computing environments.

Dr. Robert Martino heads the *Computational Science and Engineering Section*, which investigates modern, non-traditional, high-performance computer architectures for potential applicability to biomedical problems. This section also

conducts image processing as applied to chemistry and biology, operates DCRT's image processing resource, and collaborates with scientists from other ICDs in its use. A member of this section and CSL's principal image processing expert, research chemist Dr. Benes Trus, holds a joint appointment with the Laboratory of Structural Biology Research, NIAMS.

CSL initiates projects when a new computer technology emerges that potentially can benefit the NIH scientific community, such as the advanced laboratory workstation and the highly parallel computer projects. Other projects begin with requests from the ICDs for assistance in solving particular laboratory or clinical computing problems. These projects involve collaborating with a scientist, a group of scientists, an intramural laboratory, or a clinical branch who will use the systems or the methods resulting from the project. Examples include the flow cytometry advanced data analysis project, computer support for high-volume molecular biology sequencing, and investigational work into clinical image communications in the Clinical Center. Finally, CSL offers services that support distributed computing on the campus and provides advice on computer technology and laboratory computing to the intramural scientific staff, and occasionally to the extramural staff, academic institutions, and other federal agencies. To make maximum use of available resources, CSL emphasizes projects and services that potentially have wide impact on biomedical research at the NIH.

Advanced Laboratory Workstation Project Approaches Initial Deployment

In concert with the Laboratory Systems Unit of the Computer Center Branch, CSL continued the advanced laboratory workstation (ALW) project, an effort to develop a network of powerful, 32-bit, UNIX workstations, interconnected via the planned NIH campus area network, for scientists to use in their laboratories. A prominent feature of the project is a network-distributed filesystem shared by workstations throughout the campus, giving the illusion of a single, large, central filesystem; in reality, files are stored on dedicated fileservers connected to the campus network and distributed over the campus.

The decision to use the Andrew File System (AFS) developed at Carnegie-Mellon University for the ALW project received a boost this year when AFS moved into the commercial sector as a product of Transarc Corporation. The decision was further reinforced with the adoption of AFS as a part of the Open Software Foundation's Distributed Computing Environment, which may lead to widespread use of AFS later in the 1990s. Although procurement for AFS-compatible fileservers was delayed last year, benchmark testing of proposed fileservers was completed during FY90. Award of a contract for up to 16 fileservers and deployment of AFS for workstation users in many NIH buildings is expected in FY91.

To prepare for AFS deployment, CSL's Ted Persky devised a procedure for "remapping" the file systems of computers that run the UNIX operating system, a necessary step in implementing any network-distributed file system.

This procedure was applied to dozens of workstations throughout the campus and to the DCRT Convex and 3090/AIX systems. It was also made available to a number of companies and other federal agencies that used it to prepare their sites for a distributed file system.

Because the ALW project is based on portable, industry standard software technologies, CSL joined the Open Software Foundation this year, giving NIH a potential voice in the development and evolution of future software standards. As a spinoff of the software engineering work of the ALW project, Keith Gorlen and Perry Plexico of CSL joined with Sanford Orlow, an employee of CSL contractor Systex, Inc., to write a book, "Data Abstraction and Object-Oriented Programming in C++," that was published by John Wiley & Sons in June 1990.

Highly Parallel Computer Installed

CSL's plan to implement a modern highly parallel computing system as an experimental computing resource for NIH science advanced significantly with this year's installation of a Touchstone prototype parallel computer, developed under contract for the the Defense Advanced Research Projects Agency (DARPA) by Intel Corporation. The Laboratory originally had planned to initiate a procurement for a commercially available parallel computer during FY90, but fortuitously was offered the opportunity to participate in DARPA's Touchstone program.

Through its Strategic Computing Initiative, DARPA funds both university and industry research and development efforts into high-performance computing systems, then makes early copies of prototype systems available to other government agencies for use in computational science research. In the Touchstone program, DARPA contracted with Intel Corporation to develop scalable, highly parallel, high-performance computing systems. In return for access to early high-performance prototypes and gaining the benefit of DARPA's research and development funding, the collaborating agency reimburses DARPA for the basic hardware and software costs of delivered prototypes and provides performance data for DARPA to use in evaluating the programs it funds.

The Touchstone program will enable CSL to participate at the forefront of research in highly parallel computing. During FY91, CSL will begin to investigate potential applications of highly parallel computing to biomedical research, targeting such areas as computational chemistry, biomedical image processing, and molecular biology sequence analysis. Members of the CSL technical staff have been assigned for each of these three areas, and subject matter collaborators from elsewhere at NIH have been identified. Depending on the success of the initial Touchstone prototype and future developments made under the DARPA contract with Intel, CSL may elect to obtain future Touchstone prototypes or upgrades for the duration of the three-year agreement with DARPA.

Image Processing Research Advances Virus Structure Elucidation

Three-dimensional images of the herpes simplex virus type 1 were produced for the first time in FY90, as a result of CSL's applied biomedical image processing research. Collaboration with Purdue University provided the three-dimen-

sional reconstruction algorithms, which were enhanced with new software developed by CSL researchers. The images were obtained on the DCRT Image Processing Facility from digitized two-dimensional electron micrographs and have contributed to the understanding of the structure of this virus.

Interest Develops in Clinical Image Communications

Medical images are an important part of the medical record produced during a patient's hospital stay or clinic visit, and although many of these images exist in electronic form, they represent a difficult-to-manage data source because of the tremendous size of the datasets involved. The Clinical Center, like most university and research hospitals, is grappling with the problem of consolidating medical images with more conventional textual data normally managed by a medical information system.

CSL and the Clinical Center are collaborating to develop a series of demonstrations aimed at achieving image integration into the electronic medical record. Termed the Image Management and Communications System (IMACS) project, this effort will include not only conventional radiographic images and tomographic scans, but other kinds of "images" such as 12-lead diagnostic electrocardiograms as well.

This year, the project developed a plan to implement a feasibility model for integrating these kinds of images with the Clinical Center's Medical Information System (MIS). One or more of the tomographic scanners (CT and MRI in Diagnotic Radiology, and PET and SPECT in Nuclear Medicine) will be interfaced to the MIS via a commercial image gateway and a local area network connection. Because diagnostic X rays require film digitization before they can be interfaced, a procurement was initiated for a gray-scale, sheet film digitizer that can operate as an integral part of the planned image gateway. The MIS will receive electrocardiograms from the Clinical Center's existing Hewlett-Packard ECG Management System via a remote ECG workstation operating as a gateway.

Kenneth Kempner, a senior member of the CSL technical staff, was invited to participate with the American College of Radiology-National Electrical Manufacturers Association (ACR-NEMA) working group that is developing a standard for medical image communications. Mr. Kempner will represent the NIH to influence the evolution of this standard in a way that benefits the public sector.

Molecular Biology Places New Demands on Computing Resources

Advances in biotechnology during the past decade have produced major advances in molecular biology and genetic sequencing, and it has become increasingly clear that computers will play a leading role in dealing with many of the problem areas that lie ahead. Computers store previously determined molecular biology sequences, such as those in GenBank, aid in the determination of new sequences, and compare newly determined sequences for similarity to those already in a data base. As new, large-scale sequencing projects—such as those envisioned by the Human Genome Initiative—get underway, computers will play an ever-increasing role for sequence analysis, data-base accesses, and worldwide data communications.

In its first venture into the molecular biology arena, CSL began a collaboration this year with a group headed by Dr. Craig Venter of NINDS to improve computer sequence analysis technologies. Dr. Venter's laboratory operates a high-volume sequencing facility that focuses on determining the DNA sequence of the X-chromosome. It works at the leading edge of automated DNA sequence analysis, and has entered into a cooperative research and development agreement with one of the commercial leaders in DNA sequencing systems, thereby allowing it to influence the direction of this technology.

This year, CSL provided computer systems consultation and support while exploring areas where CSL and the DCRT can contribute to computer technology development in DNA sequencing. In the consultation and support areas, CSL helped implement a local area network interconnecting a variety of UNIX-based workstations, established communications between these workstations and the Macintosh-based control consoles on the automated DNA sequencers, and established a modem connection to networks on the main NIH campus as an interim measure until a planned T1 connection is installed. In addition, CSL analyzed the operations of Dr. Venter's laboratory and prepared a "working paper" identifying near and long-term areas in which further planning is needed to meet the goal of sequencing the X-chromosome.

New Support Services Complement R&D Efforts

In cooperation with other parts of DCRT, CSL initiated support services in FY90 to benefit users of scientific workstations, minicomputers, and other forms of scientific computing. Working with the DCRT Training Unit, CSL inaugurated a UNIX training program, funding the Training Unit core "Fundamentals of UNIX" course for NIH scientists and supplementing it with more specialized UNIX discussion groups led by CSL staff member John Knight. In the future, CSL will contribute technical seminars to the program that are relevant to scientific workstation usage.

With CSL assistance, DCRT's Network Task Group extended high-performance campus network connections to most of the Silicon Graphics molecular modeling workstations purchased for NIH scientists through a DCRT-sponsored procurement last year. Members of the ALW project staff from both CSL and CCB offered a range of services to users of these workstations, including software configuration and update, and technical assistance via both electronic mail and telephone hotline. Service to these and other UNIX-based scientific workstation users will expand in the coming year with deployment of the Andrew File System.

CSL began a program to offer services to users of Digital Equipment Corporation's popular VAX minicomputers that run the VMS operating system. Based on a telephone hotline staffed by CSL's technical services contractor, the program began as a six-month trial effort in Feburary 1990, and was successful enough that CSL extended it for an another six months before making a final evaluation about its continuation. Consulting assistance for VMS users on the DECnet networking protocols for communicating via computer over the campus network is also available from CSL's Dr. Ramon Tate. Dr. Tate coordinates naming and node addressing conventions for DECnet users to ensure campus-wide integrity of DECnet communications.

Future Plans

The advanced laboratory workstation project, the highly parallel computer project, and the effort to provide computer support services to scientists will continue as primary activities of the CSL in the coming year. The molecular biology venture and the clinical imaging work are expected to grow into substantial commitments. Thus, most CSL effort during the coming year will focus on continuing projects in progress, with possibly some realignment of resources to accomodate the two new projects.

During FY91, the ALW project will begin deploying AFS fileservers and extend AFS to the DCRT's Convex minisupercomputer and approximately 40 Silicon Graphics molecular modeling workstations. CSL staff members working on the ALW project will cooperate with CCB to investigate IFS, a version of AFS developed for 3090/MVS systems by the University of Michigan. CSL will evaluate and make available applications software packages of potential interest for NIH investigators that run on UNIX-based workstations, and will investigate graphical user interface (GUI) software that builds on the standard X Window System. For example, the Laboratory Applications Package (LAP), previously developed by CSL, will be expanded to work via a GUI.

In addition to investigating selected applications for the highly parallel computer, CSL will seek access to additional expertise in programming parallel computer architectures through a contractor. The laboratory will expand its capability in molecular biology by completing a recruiting action and by securing, via contract, molecular biology training for staff members and other interested DCRT employees. The training program will introduce important topics in molecular biology to engineers and computer scientists with little previous exposure to biology.

Finally, because CSL has substantially expanded its commitments to conduct new initiatives and provide services during the last year, in-house resources will be supplemented by a new technical services contract that considerably expands upon the scope and level of effort provided under previous contracts.

Research Projects

Advanced Laboratory Workstation Project

Keith Gorlen
with P. Plexico, R. Dew, T. Ghebeles, R. Malick, E. Persky, J. Powell, J. Sullivan (DCRT/CSL); J. Dickson, S. Bailey, G. Charrick, S. Fellini, J. Small (DCRT/CCB); N. Hamsath, P. Johnson, S. Orlow (Systex, Inc.)

The Distributed Systems Section (DSS) of the Computer Systems Laboratory and the Laboratory Systems Unit (LSU) of the Computer Center Branch worked jointly on a project to support researchers with 32-bit, UNIX-based, high-performance workstations manufactured by a variety of vendors. The project was based on portable, industry-standard technologies, including the UNIX operating system, the X Window graphical user interface system, the C and C++ programming languages, and the Andrew File System (AFS). Workstations will be interconnected by the NIH campus-wide LAN in order to share resources and access services such as file backup and archiving, software maintenance, applications software, online documentation, nationwide electronic mail and news, computation and database servers, laser printers, and a national distributed file system.

Applications for Advanced Laboratory Workstations (ALWs) include molecular graphics and modeling, medical image processing, gel analysis, DNA and protein sequencing and searching, statistical analysis, laboratory data acquisition, and desktop publishing.

Activities of the ALW Project in FY90

Using the Andrew File System (AFS) as a means of accessing the same files by a variety of UNIX workstation clients from different vendors, the project established two different groups of clients, or AFS "cells." A "production cell" served the needs of the ALW staff for mail, news, and demonstrations, as well as software development, configuration, and testing. A complete file backup system developed by ALW staff in the C++ language was run daily to maintain system reliability. The "test cell" allowed testing of the latest AFS software version available. It currently runs AFS Release 3.0 on a number of workstation models from IBM, DEC, and Sun. Experience with AFS 3.0 was quite favorable—it runs on the latest model workstations, is reliable, performs well, and is easy to administer.

To prepare for the deployment of the test cell software as a production-level product for NIH scientists, the CSL issued and received responses to an RFP to purchase up to 16 AFS file server machines and peripherals over the next three years. When deployed, these servers will provide over 50 GB of managed, shareable, network-accessible disk storage to NIH researchers with UNIX workstations. The RFP required each of the bidders to benchmark one of its proposed file servers on-site at DCRT.

In October 1989, about 40 Silicon Graphics, Inc. (SGI) workstations with molecular modeling software from Polygen were installed on the NIH campus. Because these workstations represented an excellent opportunity for applying ALW technology to an important area of research at NIH, the ALW staff provided UNIX systems software support for them. DCRT's Network Task Group staff worked with the ALW staff to interconnect the SGI machines via a network, achieving connectivity to machines in Buildings 2, 4, 6, 8, 10, 29, 36, 37, and 38A (approximately two-thirds of all the machines). These machines now use the Network File System (NFS) to share files, but AFS 3.0 will be made available to them by the ALW project. The ALW staff maintained telephone and electronic mail hotlines to answer questions and provide scientists with assistance in resolving UNIX problems.

For all UNIX users on the NIH campus to share a single file system such as AFS, each of them must have a unique UNIX login name, along with an integer ID which provides a unique identification to the operating system. If UNIX were new to NIH, this would be relatively easy to set up; however, the many UNIX systems already in place on the campus contain thousands of files owned by users with either duplicate login names or duplicate IDs. To enable distributed file system use, these existing file systems must be "remapped," so that each user is uniquely identifiable by both login name and number. ALW staff developed remap software and applied it to all the DCRT UNIX systems, as well as all SGI workstations with network connectivity. The software was shared with other government agencies and with private industry whose UNIX systems had similar remapping needs. In addition, the program began a central-site registry service for UNIX login names and numbers, to maintain future consistency.

This year, CSL joined the Open Software Foundation, giving NIH a voice in the development and evaluation of future software standards. ALW staffers continued work to improve the NIH Class Library, the runtime support library for the C++ Language which they initiated in 1985. Release 3.0 of the library made it compatible with AT&T's 2.1 release of the language. Several staff members completed a book about C++ and the NIH Class Library entitled "Data Abstraction and Object-Oriented Programming in C++," which was published in June by John Wiley & Sons.

Work continued this year on the NIDDK Pilot Project started in 1987 as a field test site for ALW technology. Support for the high-speed T1 link which connects Building 2 to the DCRT Ethernet passed from the ALW staff to the CCB, but CSL personnel continued to assist building scientists who serve as a Beta test site for the software packages of the ALW Project, including the Laboratory Analysis Package (LAP), X Windows, and Frame Maker. Building 2 UNIX workstations use NFS at this time, but they will be converted to also run AFS 3.0 during FY91.

Work will continue in the coming fiscal year to make every ALW workstation run the AFS 3.0 software. Staff will convert the production cell to the 3.0 software, and deployment will begin for the new AFS fileservers being obtained under contract, making AFS available to NIH scientists. The general availability of AFS will also allow for expansion of UNIX-related support services begun this year. Steps will be taken to extend AFS to the DCRT Convex minisupercomputer and to the SGI molecular modeling workstations so that users of these systems can participate in the AFS network.

Highly Parallel Computer System

R.L. Martino, Ph.D.
with C.A. Johnson, T.K. Yap, E.B. Suh, J.C. Pfeifer (DCRT/CSL)

The Highly Parallel Computer System Project will provide a parallel computer facility that will benefit the NIH staff in their scientific computing needs and develop expertise in methods of parallel processing. This system will perform computationally intensive tasks encountered in biomedical computing in much less time than conventional machines and allow the DCRT staff to master the techniques of parallel processing, which DCRT believes is an emerging computing discipline that will have significant application in biomedicine. Initial application areas that are being investigated for this new system include:

- computational chemistry
- protein and nucleic acid sequence analysis
- biomedical image processing
- biomedical signal processing
- general scientific computing.

NIH scientists will access the Highly Parallel Computer System from UNIX-based workstations connected to a campus local area network. This network connection will be accomplished with a host computer that connects the parallel system to the network, provides the software for developing applications for the parallel machine, and schedules the use of the system. This configuration will allow scientists to send their computationally intensive tasks to the parallel system from computer workstations in their laboratories, eventually making this system a computer resource for the entire campus.

This year, DCRT entered into an interagency agreement with the Information Science and

Technology Office of the Defense Advanced Research Project Agency (DARPA) in order to participate in their High Performance Computing effort. In this collaboration, DARPA provided NIH with a Touchstone prototype to develop advanced computing methods for selected biomedical research projects. Participation in the Touchstone project will enable DCRT to remain at the forefront of national research initiatives in parallel computing.

The Highly Parallel Computer System obtained from DARPA is a multiple processor system containing sixteen 64-bit processors. The primary memory is distributed among the processors where each processor has its own address space with 16 megabytes of storage. It is a multiple instruction stream, multiple data stream (MIMD) system where each processor can independently execute its own instructions on data located in its own primary memory.

In FY91, the project will begin to implement and test algorithms and software for the application areas identified above. Staff resources will be supplemented by expert parallel computer programming assistance obtained under contract. Assuming the results of early software development and testing prove favorable, the system will be expanded from 16 to 64 (and subsequently) to 128 processors.

CSL Support for High Volume DNA Sequencing

J.I. Powell
with M. Shapiro (DCRT/LSM); A. Kerlavage, M. Dubnick (NINDS/LMCN)

This year CSL began a collaboration with Dr. Craig Venter, a leader in genetic sequencing technology at NIH. State-of-the-art equipment, much of it computer-based, was already being used by the NINDS Laboratory of Molecular and Cellular Neurobiology (LMCN), Receptor Biochemistry and Molecular Biology Section, which Dr. Venter heads. In a new, mutually beneficial collaboration, CSL provided immediate assistance in integrating the laboratory acquisition and processing computer systems using available network technology, and began identifying software and hardware requirements for processing and archiving the extremely large volume of data that will be generated in the near future.

Dr. Venter's laboratory research involves sequencing large regions of the human chromosomes. A cooperative research and development agreement (CRADA) with Applied Biosystems Incorporated (ABI), one of the commercial leaders in DNA sequencing systems, provided access to the latest technology and allowed the laboratory to influence the direction of this technology. ABI and Dr. Venter's group worked on techniques to expand the sequencer's capability per gel from 16 lanes and 500 bases/lane to 24 lanes and 550 bases/lane (i.e., from 8000 bases/gel to 13,200 bases/gel). Dr. Venter's laboratory also tested an ABI prototype robotics device for automating sequence reactions.

A four-channel "gel image" from the sequencer is collected on the Macintosh II system, where initial sequence analysis is performed. "Lanes" are chosen from the image and four chromatograms per lane extracted. A proprietary algorithm is applied to the chromatograms which produces a DNA sequence (A,C,G,T or N if the base is not determined) for the lane. A sequence editor on the Macintosh II allows preliminary editing. Unlike previous editors, it displays both the chromatograms and the sequences during the editing process. The

sequences are then transferred to a SUN-4 system where they are further processed and assembled. Several different assembly packages are being explored: the Baylor NBIR package, the Wisconsin GWC package, the IntelliGenetics package, assembly on the Macintosh II using an ABI supplied algorithm, and an ABI system for the SUN-4 which incorporates a recently developed TRW chip called the Fast Data Finder (FDF). The full assembly process often involves reanalysis of the chromatograms for further editing.

This year CSL provided consultation and systems support while exploring areas where CSL and DCRT can make significant contributions. CSL assisted Dr. Venter's laboratory in setting up a sub-netted Class B Ethernet with two SUN-4/330's, a Silicon Graphics 70/4D and a Silicon Graphics Personal Iris. NFS and Yellow Pages were installed on all machines and a temporary NFS mounting scheme was set up to allow maximum usage of the limited disk capacity. The X Window System was installed on the SUNs. Hardware and software were set up so that data files may be transferred between the Macintoshes and the UNIX machines. The Berkeley print spooler, Kermit, and Adobe Transcript software were installed on the Silicon Graphics workstations. An AT&T 9600 baud modem provided an interim communications link until a T1 link could be installed. Several members of CSL visited the Los Alamos National Laboratory (LANL) to see and discuss Lab Notebook, a comprehensive data base for recording laboratory activities in molecular biology experiments used by the Human Genome group at LANL for a physical chromosome mapping project.

CSL prepared a "working paper" describing the current operations and both near- and long-range plans needed to achieve the goal of sequencing the human chromosomes. The LMCN sequencing laboratory methods are dynamic as the laboratory evaluates and tests recently developed commercial products. Major technological advances are expected to continue in this area; and while long-term goals are stable, the means to achieve them are as dynamic as the activity within the laboratory. New approaches and equipment that prove feasible will be incorporated into the production environment. As the means to achieve the near-term goals are identified, additional staff will be assigned to the project as required.

Image Management and Communication System (IMACS)

K.M. Kempner
with H.G. Ostrow (DCRT/NTG); T.L. Lewis, M.D. (CC/DIR); J.L. Foy, M.D., Ph.D. (CC/IS); G.P. McMahon (CC/DR)

DCRT and the CC began a collaboration to develop a series of demonstration projects on integration of medical images into the electronic medical record. The goal is development of a comprehensive electronic medical record that incorporates medical images with the more conventional medical data. The images involved range in size from diagnostic electrocardiograms (up to 14 Kbytes) through tomographic scans (up to 384 Kbytes) to plain X rays (up to 8 Mbytes).

Standard 12-lead diagnostic electrocardiograms were automatically acquired, interpreted, and stored on magnetic disks utilizing a Hewlett Packard ECG Management System located within the Clinical Center. To transfer ECG waveforms and their related diagnoses from this

minicomputer system to the CC Medical Information System, a remote ECG workstation was utilized as a gateway between the two systems utilizing a RS-232 pathway. Because ECG waveforms are essentially a binary image (black waveforms on white background) and because the number of equivalent black pixels in such an image is extremely low (approximately 0.1 percent), the ECG waveform data was stored and transmitted as a time-ordered list of 12-bit ECG amplitudes, rather than as a 2.75K x 3K pixel image.

Tomographic scans are produced in the Diagnostic Radiology Department on CT and MRI scanners and in the Nuclear Medicine Department on PET and SPECT scanners. In order to develop a feasibility model for integrating tomographic scans with the Clinical Center Medical Information System, one scanner will be selected from the available devices to interface via a commercial image gateway over an Ethernet pathway. It has yet to be determined which communication protocols will be utilized.

Chest X rays are routinely obtained within the Diagnostic Radiology Department,and these images may be integrated into the TDS Medical Information System or transmitted to the relevant Ambulatory Care Research Facility clinic where the patient is seen. A sheet-film digitizer interfaced to the image gateway would handle this function; therefore, a gray-scale digitizer will be acquired as an integral part of the proposed image gateway.

During the demonstration phase, images of all types that are transmitted to the Medical Information System will be stored on magnetic disks in a circular file to provide temporary working storage for the most recent images.

Viral and Bacterial Structure As Determined by Image Processing of Electron Micrographs

B.L. Trus, Ph.D.
with A.C. Steven, Ph.D., F. Booy (NIAMS/LSBR); M. Unser, Ph.D. (NCRR/BEIP); M. Kessel (NIAMS/LPB); T. Baker (Purdue University); W. Newcomb, J. Brown (University of Virginia)

Software is being developed (with NIAMS and BEIP) to analyze viral and bacterial images and perform various statistical and mathematical tests. Such images are typically analyzed by Fourier filtering techniques, the use of correlation alignment together with correlation averaging, and three-dimensional Fourier reconstruction techniques.

The electron micrographs were taken with a Philips EM400T microscope, a Zeiss 902 microscope, and the Brookhaven STEM. Some micrographs were preselected by optical diffraction. Negatives were digitized on a Perkin-Elmer 1010MG microdensitometer and analyzed by means of the PIC computer system. Results were photowritten on the Perkin-Elmer microdensitometer and on a Matrix camera station. Image processing used software developed primarily at NIH.

Study of the cell wall of *Bordetella pertussis* was continued in order to understand the structure and functioning of this cell surface. Tilt sections were analyzed in order to extend understanding of the porin channel's structure to three dimensions. Results of processing images of the sigma-1 protein from reovirus and comparisons of the images to the amino acid sequences were recently published. This protein is responsible as the receptor-recognition site for reovirus.

A new study was begun this year concerning the three-dimensional analysis of spherical viruses which exhibit icosahedral symmetry.

Using two-dimensional electron micrographs obtained for nucleocapsids of herpes simplex virus type 1, the three-dimensional reconstruction of full (DNA containing) nucleocapsids as well as empty (no DNA) nucleocapsids was completed.

Images were obtained of frozen-hydrated specimens in collaboration with NIAMS and the University of Virginia, Charlottesville, preprocessed using computer software developed in DCRT, and then processed using three-dimensional reconstruction algorithms provided as part of a collaboration with Purdue University, as well as new software developed at NIH specifically for this problem. The computational effort, using a mid-range VAX computer, was substantial, requiring approximately three to four cpu-weeks per reconstruction.

Image processing of the herpes nucleocapsids consisted of the following steps. First, individual capsids were extracted from a field of view, preprocessed to remove any background gradient, and centered in the array. Second, a brute force search of the Fourier transform of the two-dimensional image was used to locate the three-dimensional orientation, based on icosahedral symmetry and projection of the symmetry to two dimensions (the common line method). Third, solutions to the orientation for each capsid were tried in a least squares procedure on all "known" capsids, and these were refined until self consistency. The centering of the images was refined, as this can affect the quality of the three orientation angles determined. Finally, after a suitable number of images had been refined, a three-dimensional solid (e.g. 163 x 163 x 163 for herpes virus) could be viewed with surface rendering software.

Analysis of herpes family viruses, especially chemically treated nucleocapsids, will continue in order to understand the structure of the protein shell better. By removing proteins chemically and seeing the change in structure, the location of individual proteins on the surface can be clearly identified, and structure/function relationships can begin to be established. In addition, these methods may be extended to other spherical viruses which exhibit icosahedral symmetry.

Other viruses and bacteria will be evaluated for suitability for examination with these methods, and this ongoing project will be continued to determine the structure of various classes of viruses.

Image Processing of Electron Micrographs

B.L. Trus, Ph.D.
with A.C. Steven, Ph.D., E. Kolsis (NIAMS/LSBR); P.M. Steinert, Ph.D. (NIAMS/LSB); M. Unser, Ph.D. (NCRR/BEIP)

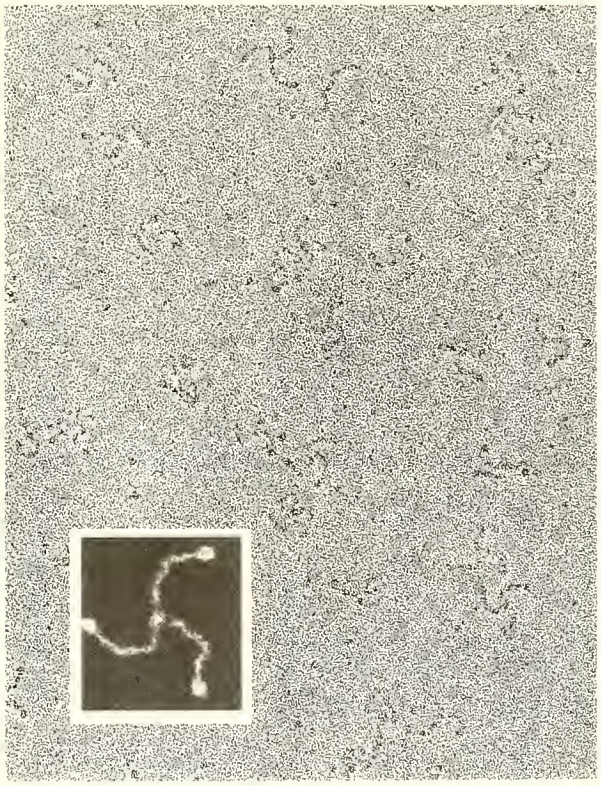
This project facilitated structure determination from electron microscopy. Suitable software, hardware, and scientific expertise was provided to allow other scientists, primarily at NIH, to use image processing and computer reconstruction to determine or understand a specimen's structure. Types of data analyzed included intermediate filaments, clathrin triskelion legs, and membrane structures.

The major results obtained this year were on the analysis of clathrin triskelion legs. Clathrin is the major protein of coated membranes involved in receptor-mediated endocytosis and other membrane trafficking reactions. The clathrin molecule, the triskelion, contains three heavy chains (190k Da) and three light chains (23-27 kDa) and assembles into a wide range of polyhedral lattices. A systematic study was

made of the structure, flexibility, and variability in leg length exhibited by isolated triskelions.

A statistical analysis of the flexibility of the triskelion legs was performed on all legs analyzed. Peaks in the curvature profile showed potential hinge regions in the molecule. Three peaks were found in the data, suggesting separate domains whose functions include both flexibility and length variability. These variations accounted for the ability of the clathrin molecule to form polymorphic forms.

Computer analysis of electron micrographs is still a relatively recent addition to the tools



Electron micrograph of purified clathrin triskelion legs, image-processed to improve signal-to-noise ratio. ***Inset:*** *one class of legs. (Image processing by DCRT and the National Institute of Arthritis and Musculoskeletal and Skin Diseases in collaboration with the University of Minnesota, Minneapolis.)*

available to scientists for structural analysis. Few laboratories have the combined software and hardware capability to perform the image processing and image reconstruction available at NIH. The techniques are especially powerful when applied to two-dimensional crystalline structures. In addition, similar particles that are not crystalline can be correlated and aligned, and the program can correct for a number of artifacts and experimental problems.

Software development will continue as needed. Some 3-D image processing capabilities may be added, and as new biological structures become available for analysis, these will be examined.

Molecular Graphics, Computer Modeling, and Sequence Analysis

B.L. Trus, Ph.D.
with S. Havlin, Ph.D. (DCRT/PSL); A.C. Steven, Ph.D., J. Conway (NIAMS/LSBR); P.M. Steinert, Ph.D. (NIAMS/LSB); M. Unser (NCRR/BEIP)

The sequence of some regular proteins, when correlated with other structural information—such as data from x-ray diffraction, fiber diffraction, electron microscopy, and spectroscopic analysis—can be used to evaluate models of protein or polymer structure. Four current studies involved the sequence analysis of keratin and other intermediate filaments (with NIAMS); computer models of biopolymers (with PSL); analysis of protein sequences from viruses and bacteria (with NIAMS); and new methods to analyze one-dimensional gel electrophoresis images (with NCRR and NIAMS).

As the complete sequence of keratin and other intermediate filaments and helical proteins becomes available, sequence analysis can proceed by studying periodicities in the sequence and by computer prediction of the

conformational properties of the specific amino acids in local regions of the chain. These predictions can be used to generalize structures where related sequences are available and to draw conclusions about similarities and differences.

Standard Fourier methods were used to analyze and cross-correlate sequences. Sequence regularities are usually correlated with structural features, such as the collagen triple helix, the alpha helix, or the tropomyosin double stranded alpha helix. Using the computer methods of Chou and Fasman or Robson, one can predict other types of conformations (e.g., alpha helix, beta sheet, beta turn). Additional software was developed at NIH to illustrate correlations, create maps of the linear sequences studied, and view inter- and intrachain homologies. More recently, multivariate statistical analysis was used to study similarity in classes of similar protein sequences, such as intermediate filaments. This research was significant because many proteins do not form three-dimensional crystalline solids whose structure can be analyzed by classical x-ray diffraction. However, if these proteins are regular, comparison and analogy with related proteins can be used to model the unknown structures in order to understand the structure and function of the proteins. In addition, computer models can be used to analyze possible protein structures based on criteria other than regular periodicities.

This year a number of analyses provided useful information into the structure of proteins. The analysis of the amino acid sequence of the sigma-1 protein from reovirus, as well as images of the proteins were recently published. Specific features in the sequence were correlated with features in images obtained from correlation alignment and correlation averaging of single fibers.

Another new analysis would using an "artificial intelligence"-type computer program to recognize alpha-helical heptads (seven amino acid segments) in amino acid sequences. The goal would be able to identify heptads automatically, and to compare the results of computer searches of sequences with human heptad searches and computer analyses based on the popular Chou-Fasman or Robson algorithms.

Another new project used computer image processing techniques and modeling-type analyses to develop Macintosh II-based software to quantitate one-dimensional gel electrophoresis. Statistical models were used to estimate background, so it may be subtracted from the intensities giving net protein amounts. A preliminary presentation was given at the industry collaboration day last fall. A few beta sites were selected to evaluate and test the new software and to participate in discussions centered on improving the utility of the operations.

As new sequences of regular (helical) proteins become available, it will be relatively easy to model these sequences and describe their structures both graphically and quantitatively.

Brain Image Registration

K.M. Kempner
with M.V. Green, M.E. Daube-Witherspoon, Ph.D., Y.C. Yan, S.D. Stein, R.G. Dextras (CC/NM); C.L. Grady, Ph.D. (NIA/LNS); J.F. Fessler (NCRR/BEIP)

An elusive problem faces researchers involved in correlating brain form (structure) from x-ray computed tomography (CT) images and brain function (metabolism) from nuclear medicine positron emission tomography (PET) images. The difficulty concerns the superposition and

registration of the tomographic views obtained from these two imaging modalities.

The driving force behind the brain image registration project is the need to develop a greater understanding of the processes underlying the generation of PET images. It is hoped that development of techniques for accurately correlating CT structural data with PET metabolic information will enhance this understanding.

Two stages were used to approach the problem. First, practical methods were developed for the accurate and reproducible placement of the head within a tomographic scanner's aperture. Second, simplified algorithms were developed for the scaling and registration of digitized images from different scanners on a digital display subsystem.

Precise orientation of the subject's skull within the scanner's aperture was monitored and recorded through the use of a Polhemus Navigation Systems positron/orientation transduction subsystem connected to a PC-XT. Two inexpensive custom-molded oral appliances were developed to allow the position/orientation subsystem's sensor to be fixed to the subject's skull. A novel targeting algorithm was derived to provide the technician operating the system with visual cues related to head position within a scanner's imaging volume.

Methodology was developed by Nuclear Medicine Department staff which allowed list mode data collection from a PET scanner to be correlated with continuous measurements from the position/orientation subsystem. This permitted PET image reconstruction to be performed while including a correction for subject motion during the scan. A similar result was achieved with a scintigraphic camera by taking serial images and using measurements from the position/orientation subsystem. Correction for subject motion in the image plane was achieved by appropriate rotation and translation of the serial images prior to summing these images. Feasibility studies utilizing these techniques were performed with phantoms under controlled conditions.

Future efforts will center on clinical testing of the accuracy and repeatability of skull placement in tomographic scanners and on refinement of algorithms for removing motion artifact during the scanning process.

Flow Cytometry Advanced Data Analysis

L. Barden
with R. Tate (DCRT/CSL); S. Sharrow (NCI/DCBD/EIB); D. Plugge (Systex, Inc.)

The Flow Cytometry Advanced Data Analysis Project is a collaborative effort between the Computer Systems Laboratory and the Experimental Immunology Branch (EIB), Division of Cancer Biology, Diagnosis and Centers, NCI. The EIB flow cytometry facility provides consultation in flow cytometry techniques, protocol design and reagent selection, as well as data acquisition and storage and analysis services to more than 55 NCI investigators. Several other flow cytometry facilities in the NIH Intramural Research Program are less directly involved but benefit from the software developed for this project. The current goal is to design and implement a state-of-the-art, production-oriented research facility to acquire, archive, and analyze multi-parameter flow cytometry data. The new design will be based in part on an existing data acquisition and analysis system developed and enhanced by EIB and CSL over the past 15 years.

Commercially available flow cytometry data analysis software is oriented primarily toward

single user applications rather than to providing production service to a large number of investigators. As the Experimental Immunology Branch facility supports multiple research projects for many investigators, it was necessary to provide a more automated (production-oriented) environment than is commercially available. To this end, CSL continued developing the Cluster Analysis Program, which provides non-hierarchical cluster analysis of multi-parameter data, and the Laboratory Analysis Package, for indepth analysis of single and dual parameter histograms.

In the latter half of this fiscal year, CSL began an extensive design and development effort to provide an automated system for archiving and retrieving data that would be driven by user (or client) program requests for storage and retrieval services without operator intervention.

The EIB flow cytometry facility was connected to the NIH-wide network and to the EIB local area network, providing direct access to experimental data and computational resources from networked laboratory personal computers.

During FY90, the following major equipment items procured with NCI FY89 funds were delivered and integrated into the EIB facility:

- a Becton-Dickenson FACSStar Plus flow cytometer equipped with a Digital Equipment Corporation VAXStation 3500 workstation for instrument control and data acquisition
- a System Industries multi-ported disk subsystem three gigabytes of on-line disk storage
- a SUMMUS magnetic tape carousel loader which stores 54 8-mm magnetic tape cartridges, providing access to 125 gigabytes of near on-line archival quality data storage
- a MicroVAX 3800 computational server to supplement the existing MicroVAX-II processor.

The development of the Cluster Analysis Program (CAP) and the Laboratory Analysis Package (LAP) continued this year. Provision was made for multiple displays per screen and multiple plots per page, a significant expansion to the written documentation and conversion of on-line documentation to be compatible with VMS Help. Ease of user interaction during the post-clustering data analysis phase was enhanced. Conditional code was inserted in the base copy of LAP (maintained under UNIX as an Advanced Laboratory Workstation application) to provide direct hosting of LAP on VMS, obviating the need to port each improved version of the program. Support for the standard flow cytometry text block was enhanced in both programs.

The initial design phase of the data archiving package and a program to automate the generation of histogram files from listmode data were completed, and the development of prototype code for both these applications was begun.

During FY91, the data archiving system and the listmode-to-histogram data reduction program will be completed, the vendor-supplied data acquisition system will be integrated into the data archiving and analysis systems, and the results will be tuned for reliable production use. Both CAP and LAP will be delivered to other flow cytometry sites, both on and off the NIH campus. PC-based data analysis of data acquired at this facility by DCBDC researchers is also expected to expand.

If resources permit, the range of available data analysis techniques will be expanded beyond simple cluster analysis and histogram

integration, and an optical disk jukebox for near on-line archival storage of data will be procured.

Laboratory Application Package

J.I. Powell

The Laboratory Analysis Package (LAP) is a computer program to analyze and display data such as that produced by spectrometric means. An extension of the previous APR, DP2 and DP3 programs developed by the Computer Systems Laboratory, it was adapted to operate in the UNIX domain. Although substantially more powerful than its predecessors, LAP remains an easy-to-learn and easy-to-use interactive program. It is written in the C++ programming language and was developed and extensively tested on a Sun-3 computer.

LAP can perform a wide range of data manipulations on vector data, x-y paired data, and matrix data, using either a command or expression syntax. Customized command procedures can be saved in files and added to the LAP command set. Results may be viewed as line graphs, scatter graphs, bar graphs, perspective views, or contours on a variety of graphics devices including those that support X Windows. Publication quality plots may be produced in several formats, including encapsulated postscript.

Major accomplishments of the LAP project this year

- New capabilities were added including contours, cubic spline fit, additions, and enhancements in the graphics to allow more flexability in producing publication-quality plots. Work was initiated to add HPGL-formatted graphics output.
- The program was ported from C++ 1.2 to C++ 2.0.
- Work was initiated on porting LAP to the NIH Convex, Sun-4's, DECstations running Ultrix, and Silicon Graphics.
- Coordination with the Advanced Data Analysis Facility Project in porting LAP to run under VAX VMS was continued.

During the next fiscal year, the ports of LAP will be completed and made available via FTP out-side NIH. The addition of more rigorous curve fitting and modeling capabilities to LAP will be explored and LAP will be made more useful and usable by the development of a graphical user interface with one of the X Window toolkits (Motif, Open Look, or Xview).

Zeiss EM902 Electron Microscope

W. Gandler
with B.L. Trus (DCRT/CSL); A.C. Steven, M.D. (NIAMS/LSBR)

Two years ago the Laboratory of Structural Biology Research, NIAMS, acquired a Zeiss EM902, the first commercially available electron microscope to incorporate an imaging energy-loss spectrometer. The main thrust of the project—and the reason for procuring one of these instruments—was to exploit its potential for imaging relatively thick specimens in vitreous ice. Preserving the native structures of biological macromolecules by rapid freezing and maintaining them in an aqueous environment (ice) during imaging has been one of the most important breakthroughs in electron microscopy of the last 20 years, and the Zeiss EM902 has considerable potential to advance this area of research.

CSL proposed a computer system based on the MicroVAX II to acquire data from the EM902. Image acquisition is accomplished via a Dage-MTI SIT-66 camera provided by Zeiss which is incorporated into the microscope.

Because the microscope is often operated under minimal dose conditions, multiple images are averaged for image intensification. Video output from the camera is fed to a "frame grabber" and image processing board (Data Translation DT 2651 and 2658) connected to the MicroVAX I/O bus. A FORTRAN acquisition program using DT-IRIS subroutines was developed. The program acquires one of every four video frames (CCIR standard), averaging them in real time in a 512 x 512 matrix. The averaged images are stored on disk in the same format used by the CSL image processing facility, allowing extensive computer image processing to be performed with the highly developed PIC software package on either the large VAX system in DCRT or locally on a MicroVAX III. Either system may be accessed by a network connection. DECnet is used to transfer images on a 56K bit/sec link to the DCRT VAX and a local area VAX cluster connection enables local image transmission.

A shading correction program was added this year to correct for the effect of uneven illumination across the object. The program requires two reference images to be collected, along with all the original uncorrected images acquired at the same intensifier voltage, video gain, and black level settings. One reference image is the dark current image, acquired with no light present; the second is a shading correction mask, which is the image obtained from a uniformly illuminated field with no specimen. When the gray scale values are a linear function of optical density, a corrected image can be calculated.

The next step in system development will be to measure geometric distortion and add a program to correct for it if required. A linearity chart (Ball chart) has been ordered to determine geometric distortion.

Computer Assisted Patient Interviewing in Clinical Pharmacy

J.M. DeLeo
with F. Pucino, K. Calis (CC/Pharmacy Department)

Clinical pharmacists are responding to the drug information explosion by participating more directly in patient care. Specifically, clinical pharmacists have become increasingly active in dispensing pharmaceutical information to patients, checking for patient compliance with their medication regimen, screening for adverse effects caused by single drugs and drug combinations, and sorting out possible complications between medications and food, allergies and chronic medical conditions. To be fully effective in these increasingly demanding tasks, pharmacists must keep abreast of new drug knowledge, allocate more time for patients, and continually sharpen their interviewing skills.

In order to explore potential roles of the computer in these emerging tasks, a collaborative project was initiated between the Computer Systems Laboratory and the Clinical Center Pharmacy Department to design and develop a computer system that could directly interview patients in order to collect medication histories and identify possible untoward effects related to medication regimens. It was believed that this "patient-friendly" system, if successful, could truly mimic the skills of clinical pharmacists and permit multiple simultaneous interviews. Not all patients are suitable candidates for direct computer interviewing. It was, however, a design objective to simplify the computer interview process so that a large number of patients could be accommodated. Other

significant objectives of this project included providing a high degree of flexibility in writing, installing and modifying new interview scripts; maintaining on-line comprehensive information on drugs and adverse effects; and making computer interviewing available in the NIH Pharmacy outpatient waiting area as well as in all of the NIH clinics supported by the Pharmacy Department.

Work has begun on designing and writing a comprehensive interview script comprising ten subscripts that collect the following information: patient description, medical conditions, medication regimen, medication compliance, symptoms and implied medications, allergies, dietary history, social history, occupational and environmental exposure, and patient evaluation of the system. These subscripts are being written in a new high-level language that was designed to facilitate interview script writing. An interpreter processing these subscripts was designed and is being written in the BASIC language. The dual interpretative/compiler nature of modern BASIC, combined with its machine independence and ubiquitousness, make it suitable for this project. Work was begun on generating a comprehensive list of lay-language adverse-effects terms, grouped by body systems. This list will be transformed into an efficient questionnaire subscript for screening patients for adverse effects in an unbiased fashion. Several commercial drug information data bases are being investigated as candidate data bases for use with the system.

The first major milestone of this new project, implementing the interview system independent of a drug data base in order to get experience with the response of patients using the system, took place in late FY90.

The CSL staff acknowledges the members of the United States Pharmacopoeial Convention and the Food and Drug Administration who provided helpful information for this project.

Evaluation of patient and pharmacist reactions to the initial testing will be done in FY91. In addition, selection and incorporation of a drug data base will be significant milestones for the coming year.

Neuromagnetometer Computer System

R.L. Martino, Ph.D.
with C.A. Johnson, T.K. Yap, J.E. Sullivan (DCRT/CSL); S. Sato, M.D., M. Balish, M.D., P.V. Connaughton (NINDS/Office of the Clinical Director)

The Medical Neurology Branch, NINDS, and the Computer Systems Laboratory, DCRT, are collaborating on a research project to localize epileptic discharge sources within the human brain noninvasively by using neuromagnetic recording in conjunction with conventional electroencephalogram (EEG) recording. Many patients with seizure disorders exhibit low-level cellular discharges between seizures, indicated by interictal spikes or sharp waves in their EEG and magnetoencephalogram (MEG) recordings. This project involved developing computer techniques for automating and enhancing the procedure presently used by NINDS neurologists to determine the intracranial locations of the sources of epileptiform discharges in patients with epilepsy.

During the past year, a CSL-designed computer system that detects epileptiform discharges from the EEG and MEG signals in real time was placed into clinical operation. The system was implemented with a number of detection algorithms—some developed by CSL

and some selected from the published literature—and a variety of options for configuring these algorithms, allowing the medical staff to choose the optimal method for a given patient. The system provides a realtime display of the signals showing where an event is detected and allows neurologists to save both epileptiform discharges and longer seizure activity manually. Clinical evaluation of the system was begun. Development of three-dimensional displays of a patient's head using the digitized outline of the the head and anatomical data obtained from computed tomography and magnetic resonance imaging scans continued, using the Mayo Foundation ANALYZE software package on a UNIX workstation. Neurologists will view these displays to determine where epileptic sources are located relative to actual brain anatomy.

In the coming year, CSL will perform a comprehensive clinical evaluation of the real-time detection system and compare its performance to that of a new commercial system that uses an algorithm not available on the CSL system. Work will continue on the development of the three-dimensional anatomical displays.

Lipid Analysis Sample Tracking System (LASTS)

R.L. Tate, Ph.D.
with J. Hoeg, M.D., D. Wood (NHLBI/MDB)

Over the past 20 years the Molecular Disease Branch (MDB) of the NHLBI has studied human lipid metabolism disorders by analyzing tens of thousands of blood samples from nearly 7,000 individuals. Until quite recently, all the accumulated data was gathered and entered entirely by hand into central NIH Computer Utility data bases. As the volume of samples and personnel workloads increased, it became increasingly difficult to process and enter data in a timely way. Furthermore, the advent of automated blood chemistry analyzers and powerful personal computer database software pointed the way to a comprehensive automated method for tracking samples and acquiring and processing the analytical results.

LASTS is a comprehensive PC-based system for recording the results of lipid analyses performed on plasma samples submitted to the Molecular Disease Branch Lipid Analysis Laboratory. Identifying information about the samples is entered into a data base maintained on a laboratory PC and verified when the sample is acquired. The samples, identified by bar-coded labels, are subdivided for analysis. As each analysis is performed, the results are either captured directly from the analyzer or keyed by the encoded label for manual entry. The combined results to date on each sample are maintained in a data base that can be searched by laboratory personnel or the referring physician. Once the test results have been verified, the sample data are copied to a report dataset that is transferred to the NIH Central Computer Utility and incorporated into the MDB lipid study data bases. Verified data is also maintained locally in a form suitable for access by PC-based data-base query programs.

Although the system design is complete, several interim programs are now in use while the final system is being implemented. Data entry is still being performed manually, and the data acquisition and verification routines have yet to be implemented. Since MDB is in the process of acquiring a more capable automated analyzer, modifications to the analysis planning and data acquisition modules will be necessary. Clinicians may now request summary information about analytical results on samples re-

ceived within the past 2 1/2 years, and a semi-automated procedure has been put in place for transferring new data from the MDB data bases to the NIH Computer Utility master data files. Completion of the remaining system components is anticipated within the next fiscal year.

Data System for Objective Analysis of Clinical Cataracts Using the Zeiss Scheimplug

M. Vivino
with B.L. Trus, Ph.D.(DCRT/CSL); M. Datiles, M.D., J. Schumer, M.D., M. Sibug, M.D., B. Magno (NEI/CB)

The Clinical Branch, NEI used images produced by the Scheimpflug principle to quantitate opacities in order to find ways to determine changes in a cataract lens accurately. The Clinical Branch employed a Zeiss Scheimpflug Slitlamp Camera (SLC) to visualize pathological changes in anterior ocular measurements and to make geometric measurements of boundary surfaces and their distances. The SLC measuring system consisted of a special photo slit lamp with a computer-based image display and analysis system which was used clinically primarily to measure cataract area and opacity.

In past years, CSL responded to a need in the NEI/CB by replacing and enhancing the data system for the Zeiss SLC. The camera and an image acquisition and processing system were delivered to the National Eye Institute in 1986; support for this original system was reported in the 1987 DCRT Annual Report. A primary system component which showed some degraded system performance was the analog video recorder used for storage and recall of images.

This year, CSL helped NEI develop a Macintosh II computer-based system for image acquisition and processing and used an optical disk for image storage and retrieval. The original instrument utilized an elaborate feedback system for control of light intensity and synchronization, which was retained. The existing Bosch Newvicon Video Camera was connected to Data Translation's Macintosh II-compatible Quick Capture Image Acquisition System. Quick Capture includes an 8-bit video ADC operating at 25 frames/second, and writes data to a 768 x 512 vertical image buffer.

IMAGE, a program for general-purpose image analysis on Macintosh computers was selected for use in this application. Developed by Wayne Rasband at NIMH, IMAGE was modified to respond to external synchronization. IMAGE can calculate intensity profiles, histograms, and mean values within user-selected regions and can also export pixel values to a graphics package.

CSL installed the image acquisition hardware, designed and built a circuit to synchronize the Macintosh II to external events, and installed and modified the software as required. A data base was installed on the Macintosh to maintain patient records. In addition, software was developed on a VAX computer system to assist in detecting, quantitfying, and classifying cataracts. Successful in showing the differences between different subclasses of cataracts, the software may be useful in longitudinal studies involving anti-cataract drugs.

In the coming year, CSL will collaborate with NEI/CB in developing of new methods to improve selective imaging of cataracts *in vivo*. In progress is a study to determine the value of different wavelengths to enhance cataract imaging. CSL also is developing an addition to the SLC which will enable a physician to study how a patient with an artificial lens can focus. The addition will provide near- and far-fixation

targets on which a patient will focus while the SLC captures an image.

Cataract Data Storage and Retrieval System
H.A. Fredrickson
with J.E. Lee, M.B. Datiles, M.D. (NEI/CB)

The system provided support for an existing IBM PC dBase III+ file structure of cataract patient data in the Clinical Branch, NEI. Supplemental general-purpose programs were developed to manipulate dBase III+ files with greater ease, performance and added capability. A KWIC Index program (Key Work In Context) was designed for processing non-numeric data (free format text), to allow users with lengthy text fields in their files to perform fast interactive searches on any "Key Words" in these fields.

CSL will evaluate the general-purpose software developed this year, to determine its broader usefulness to the NIH.

100-Channel High Speed Spectrophotometer
H.A. Fredrickson
with A.R. Schultz, (DCRT/CSL); W. Friauf, J. Cole, P. Smith, Ph.D. (NCRR/BEIP); R.W. Hendler, Ph.D. (NHLBI/LCB)

A computer-controlled 100-Channel High-Speed Spectrophotometer is being developed by the Biomedical Engineering and Instrumentation Program of the National Center for Research Resources and CSL for the Laboratory of Cell Biology, NHLBI. It will be used to obtain more complete spectral information concerning the rapid changes of the reduction and oxidation centers within the protein enzyme cytochrome oxidase. This enzyme is involved in cellular respiration and is located within the inner lipid bilayer of the mitochondria.

The electronic hardware will consist of two 50-element photodiode arrays connected to 100 discrete A/D converter and local storage channels. Each channel is capable of acquiring data every 10 microseconds. A DOS-based computer will be used to control the spectrophotometer. Timing control data will be transmitted from the PC, and the PC will receive data from the 100 A/D channels via a 40-bit parallel interface.

The Computer Systems Laboratory will assist in interfacing the spectrophotometer to the PC and will develop the data acquisition and control software. It is anticipated that the MLAB software, originally developed at DCRT, will be used for the data analysis.

The spectrophotometer hardware is still in development, with usable data from the unit expected within the current year.

Auditory Brainstem Response (ABR) Analysis and Interpretation Expert System
J.M. DeLeo
with A. Pikus, D.A. Sklare (NIDCD)

The objective of this project was to build an expert system to analyze and interpret the Auditory Brainstem Response (ABR), an electrophysiological response of the human brainstem to acoustic stimuli. Expert reading of the ABR is essential in clinical decisions concerning retrocochlear disorders.

A rule-based expert system was developed to perform neurodiagnostic interpretation of the auditory brainstem response. Developed on the PC with the commercial expert system shell GURU, the system has been operational in the NIH Audiology Clinic for over one year.

ABR interpretation strategies differ considerably among various clinics. This system is a tool for rapidly modeling any ABR interpretative paradigm. Paradigms are easily added, modified, deleted, and selectively engaged at run

time. Paradigm rules are boolean logic expressions containing discrete, continuous, and fuzzy variables representing ABR, subject, and stimulus parameters. The system models human ABR expertise in explicit representational form, handles routine cases well, reduces diagnostic errors of omission, and facilitates understanding of ABR clinical decisionmaking. It can be used as a methodology for testing and validating any ABR interpretative strategy, a repository for meaningful ABR interpretative expertise, and a medium for sharing ABR knowledge.

Experience in developing and using the system suggested the possibility of generating more accurate ABR neurodiagnostic interpretation paradigms directly from ABR data using machine-learning methods. A study to derive diagnostic expert system rules directly from ABR data was therefore initiated. A significant awareness arising from this project was the importance of two separate methodologies, namely multi-valued logic (fuzzy logic) and the Receiver Operating Characteristic Curve (ROC), in modeling ABR clinical decisionmaking. Fuzzy logic has provided a mathematical structure for constructing hypotheses (rules) that combine hard, quantitative variables (such as wave latencies) and soft, qualitative variables (such as wave morphology). The ROC function has provided a tool for specifying more generalized clinical decision rules as well as a way of measuring the accuracy of these rules in terms of sensitivity (true positive fraction) and specificity (true negative fraction). Traditionally, ROC methodology has been developed for binary event data, e.g,. absent or present, negative or positive, diseased or normal. There are many conceptual entities in clinical medicine which are not easily described by a two-state model, such as fever, obesity and left ventricular hypertrophy. In such cases, fuzzy logic offers a more suitable framework. Fuzzy logic does not replace binary logic, rather it extends it. Two-state logic is a special case of multi-valued logic.

The need to assess the accuracy of the multivalue clinical decision rules that contain fuzzy variables has been recognized. Traditional ROC methodology, which is well established in evaluating accuracy in two-state models, was extended to include fuzzy models in a recent publication by Campbell and DeLeo entitled, "Fundamentals of Fuzzy Receiver Operating Characteristic (ROC) Functions." Performance of the fuzzy ROC approach, using the area under the ROC function (maximum sensitivity plus specificity) was considered. Discussions concerning the application of fuzzy ROC methodology to ABR work and other clinical and basic biomedical research areas are ongoing.

DECnet Management

R. Tate, Ph.D.

The campus-wide networks were expanded this past year to include over 75 systems using the DECnet networking protocol. Coordination of this expansion required careful planning and cooperation among the ICDs involved. The principal investigator for this project is providing the centralized coordination necessary for smooth incorporation of new DECnet hosts into an integrated network that spans both the NUnet and the RESnet campus networks. It is estimated that extension of network services to all NIH buildings will increase the number of interconnected DECnet nodes to over 150 within the next year.

UNIX Training
J. Knight

The UNIX Training Project helped both inexperienced and sophisticated computer users to become comfortable, productive UNIX users, while studying and reporting on the technology transfer issues encountered and solutions found.

The project strove to manage training activities to suit the requirements and schedules of NIH researchers.

In FY90, the project team:

- Assisted CCB in evaluating a two-day "Fundamentals of UNIX" course. The course was offered once a month and provided researchers a baseline of UNIX skill.
- Established regular walk-in UNIX discussion groups on an introductory and advanced level every Monday and Thursday afternoon.
- Developed a one-hour presentation, given on the first Tuesday of every month, to describe UNIX documentation and emphasize hands-on use of online documentation.
- Evaluated and made available to researchers videotaped courses, online courses, and courses based on special devices.
- Started a newsletter to provide information on UNIX-related events.

In FY91, the project team plans to:

- Use data collected in the discussion groups to discover ways to match training and researchers needs.
- Develop training programs to help researchers benefit from the power of emacs, X Window System, Andrew File System and UNIX communications facilities.
- Assist NIH researchers to build a working community on the UNIX platform.

Medical Information Technology Project
S.I. Allen
with C.S. Brown, M.D. (Bethesda)

This project was designed to develop better ways to let physicians and their associates use computers in health card recordkeeping for research and patient care. The methodology focuses on providing disease-specific and problem-specific protocols and hierarchies of information that allow rapid convergence on relevant diagnoses, treatments, and procedures.

In past years, computer programs were developed for the physician to produce pharmacy prescriptions and drug-related patient information using high-speed menu selection methods. Later, new modules to aid in producing diagnostic schedules and treatment reports were developed. All these programs run on a personal computer (PC).

Pilot studies with a dermatologist (C.S. Brown) involved direct entry of medical transactions into a computer terminal by the physician and staff during the doctor-patient encounter. Initial results with the PC system showed more precise and more rapid prescription-writing and procedure reporting. More recent experience demonstrated the ability of the practicing physician to organize and maintain a personal patient-care data base of drugs, tests, reports, and progress notes.

This year, the system was converted to Macintosh hardware to exploit its user-friendliness, graphic features and mouse-directed control. A conventional Macintosh menu interface was developed in the object-oriented language SuperCard, a superset of HyperCard.

Programmer-specific edit and utility pulldown menus provided data-base definition, update, and consistency checks; user-specific popout menus and graphic selection techniques controlled patient document composition and output. These studies were presented at the 14th Symposium on Computer Applications in Medical Care (SCAMC) in November 1990.

Publications and Presentations

Aldroubi, A.; Unser, M.; Tietz, D.; Trus, B.: Computerized methods for analyzing two-dimensional agarose gel electropherograms. *Electrophoresis* (in press).

Campbell, G.; DeLeo, J.M.: Fundamentals of fuzzy receiver operating characteristic (ROC) functions. In: Berk, K., (ed.) *Proceedings of the Twenty-first Symposium on the Interface, Computing Science and Statistics*, Los Alamitos, California 1989.

Douglas, M.A.; Trus, B.L.: An introduction to image processing in medical microscopy. *Medical Progress through Technology* 15, 109-140 1989.

Edwards, P.A.; Datiles, M.B.; Unser, M.; Trus, B.L.; Freidlin, V.; Kashima, K.: Computerized cataract detection and classification. *Current Eye Res.* 9(6), 517-524 1990.

Fraser, R.B.; Furlong, D.B.; Trus, B.L.; Nibert, M.L.; Fields, B.N.; Steven, A.C.: Molecular structure of the cell-attachment protein of reovirus: correlation of computer-processed electron micrographs with sequence-based predictions. *Virology* 64, 2990-3000 1990.

Gorlen, K.E.; Orlow, S.M.; Plexico, P.S.: Data abstraction and object-oriented programming in C++. London: John Wiley & Sons 1990.

Johnson, C.A.; Martino, R.L.; Yap, T.K.: Real-time detection of epileptogenic discharges using autoregressive prediction. *Proceedings of the Sixteenth Annual Northeast Bioengineering Conference, Institute of Electrical and Electronic Engineers,* New York, pp. 89-90 1990.

Kashima, K.; Datiles, M.B.; Trus, B.L.; Edwards, P.A.; Kaiser-Kupfer, M.I.: Localization of lens opacities in gyrate atrophy: image analysis study with Scheimpflug camera. In: *Folia Ophthalmol. Japonica* pp. 222-224 1989.

Kocsis, E.; Trus, B.L.; Steer, C.J.; Bisher, M.E.; Steven, A.C.: The clathrin triskelion leg has an extensible proximal domain. In: *Proceedings of the XIIth International Congress for Electron Microscopy* pp. 468-469, San Francisco, California 1990.

Persky, E.M.: Features of the NIH class library. The C++ Report 2, 8 July 1990.

Unser, M.; Trus, B.L.; Eden, M.: Iterative restoration of noisy elastically distorted periodic images. *Signal Processing* 17, 191-200 1989.

Unser, M.; Trus, B.L.; Frank, J.; Steven, A.C.: The spectral signal-to-noise ratio resolution criterion: computational efficiency and statistical precision. *Ultramicroscopy* 30, 429-434 1989.

Unser, M.; Trus, B.L.; Steven, A.C.: Normalization procedures and factorial representations for classification of correlation-aligned images: a comparative study. *Ultramicroscopy* 30, 299-310 1989.

LAS

Laboratory of Applied Studies

John E. Fletcher, Ph.D., Chief

The Laboratory of Applied Studies (LAS) is a multidisciplinary laboratory staffed by mathematicians, computer scientists, physicians, and engineers. Operating in a task-oriented mode, the LAS develops solutions to research problems by examining the underlying scientific principles, identifying the appropriate mathematical and engineering concepts, and utilizing computing systems to carry out the research objectives. Many LAS projects are collaborative efforts with laboratory scientists and clinical investigators at NIH or at other research centers.

LAS investigations range from direct participation in clinical and laboratory activities to abstract development of mathematical theories and methods and computer algorithms essential to quantitative computer modeling. Software and analysis techniques developed in such investigations are then made available as general research tools. The activities are carried out by the *Medical Applications Section*, staffed by physician-scientists, electronics engineers, and computer systems analysts; and the *Applied Mathematics Section*, whose staff includes specialists in applied mathematics, computer science, and computer modeling. LAS projects may also involve guest workers, university students or other volunteers.

Intramural Research Support and Collaboration

LAS activities in FY90 included the continuing transition and expansion of computing systems to PC's and networked distributed computing as well as a broad involvement in the NIH intramural research program. Desktop computing continued to be investigated throughout the laboratory in support of both administrative and scientific computing. Two MacIntosh systems, MAC IIs, were installed to broaden the experience of the scientific and administrative staff with alternative approaches to computing. The shift from mainframe scientific text preparation to newer PC-based on-screen systems was completed this year. A fully expanded PS/2 model 80 was installed as a mathematical software test bed, and full-scale tests of CSI-PC-MLAB, MATHEMATICA, and MATLAB were carried out. Initial results indicate that such systems are promising and they may become primary systems for scientific modeling. The installation of laser printers to support many of scientific PC configurations considerably enhanced manuscript and graphics printing capability.

Applied Mathematics Section

Many of the numerical analysis, mathematical modeling and data analysis systems available at NIH were adapted for operation on desktop systems during FY90. Courses were designed and offered to familiarize users with these modeling systems. Demand for such training is expected to increase as the CONVEX and other UNIX-based systems come into the mainstream of scientific computing and the modeling systems become available on these machines; therefore, training courses may be repeated periodically.

General purpose software for modeling reaction-diffusion systems in one space dimension was developed during FY90 and a course in its use was offered. The programs were considered for the addition of graphics capability, but working memory limitations have prohibited its use. Fine tuning of the software to improve user

friendliness and portability on desktop systems, as well as consideration of visual enhancements are planned for the coming fiscal year.

This software and other LAS-supported mathematical tools and software are integral parts of ongoing studies on

- monoclonal antibody function (with scientists in NCI)
- diffusion and transport in micro-dialysis (DRS)
- cross-link function in muscle dynamics (NIDDK)
- hemoglobin-oxygen binding configurations (LAIR)
- thermal unfolding of proteins (NIDDK)
- ATP-ADP-PO4 system reactions (NHLBI)
- redox states of iron in cytochrome oxidase (NHLBI)
- T-cell reactions in HIV infections (NIAID)
- NMR spectra of antitumor drugs (NCI)
- analysis of ricin dissociation (DRS)
- internal research projects that involve DCRT/LSM and PSL.

Medical Applications Section

This section continued its investigations into the prognostic power of the ECG, signal-processing methods in electrocardiology, filter design for ECG analysis, and nonbiased methods for baseline wander removal in ECG analysis continued in FY90.

Strong support of image processing and image analysis was continued in collaboration with Clinical Center and Nuclear Medicine projects. The MIRAGE system was ported to several other NIH computer systems and has been expanded to include three-dimensional capability. Consultation and collaboration was provided for PC- and Macintosh-based imaging systems and specifications were developed for a broad-based radiation science imaging system. The LAS was also involved in the design and specification of a Clinical Center image transmission and archiving network.

The system design for the Computer-based Monitoring of Central Nervous System Function in the Neurosurgical Intensive Care Unit (Cleveland Clinic) was completed and published. A technical survey of computer algorithms to peruse molecular data bases and analyze macromolecular sequences was completed this year and a resulting manuscript was accepted for publication.

Staff Activities

J. Fletcher (Chief, LAS) continued to serve as chairman of the Mathematics and Computer Science Department for the Foundation for Advanced Education in the Sciences (FAES) and as consultant to NIH intramural scientists on mathematical methods and software for ordinary and partial differential equations. He served as an adviser to the National Counsel on the Teaching of Mathematics (NCTM) commission for mathematics curriculum standards and served as a graduate thesis committee member for Southeastern University.

R. Shrager (AMS, LAS) continued to serve as consultant to NIH intramural scientists in areas of data analysis and experimental design, particularly in areas involving model-fitting and digital filtering.

J. Bailey (Chief, MAS) led a group in establishing guidelines for the practice of automated electrocardiography, which have been adopted by the American Heart Association and pub-

lished. He served as a member of the Association for Advancement of Medical Instrumentation (AAMI) ECG Committee, Secretary-Treasurer of the International Society of Computerized Electrocardiology (ISCE), and as a consultant to the European Economic Community program on "Common Standards for Quantitative Electrocardiography." He currently leads a subcommittee of the AAMI which is developing standards for ambulatory electrocardiographical devices.

M. Douglas (MAS, LAS) is chairperson of the Women's Advisory Committee. She managed the logistics for Career Day; performed analysis of workforce data; served on the selection review board to evaluate candidates for the Federal Women's Program Manager position; and was instrumental in founding the NIH Day Care Committee. She also served as the DCRT representative to the NIH Advisory Committee for Women's Health Issues.

Future Tools, Trends

Support will be continued for the MATLAB, MLAB, ML/e, and MATHEMATICA mathematical modeling systems. Where feasible, the modeling and data analysis systems will be fully implemented on both the mainframe and distributed computing systems and PC's with connections via networks. Numerical software for solving Partial Differential Equations will be investigated to include two space dimensions, and new applications will be considered in areas of microtubular perfusion, monoclonal antibodies, and in systems related to muscle mechanics. Ongoing collaboration projects with intramural scientists will be continued wherever appropriate.

Research Projects

The Solution of Reaction Diffusion Systems in Biology

John E. Fletcher, Ph.D.
with G. Weiss (DCRT/PSL); J. Weinstein (NCI/DCBD)

This project is concerned with numerical methods and mathematical software for the solution of ordinary and partial differential equations that describe dynamic physiological processes.

During FY90, software to solve single and coupled systems of reaction-diffusion equations was completed and tested on the IBM 3090 mainframe, the PC-AT, PS 2/60, and the PS 2/80. All programs were developed in FORTRAN77. The PDECOL system was revised to use the same input files, adapted for general-purpose applications, and tested on the same machines. A course detailing the use of this software was presented in April 1990 and copies of the software were distributed.

Future Tools, Trends

Future work will concentrate on exploring applications of the available software on a variety of personal computers and consider the integration of visual display routines into the packages. Some extensions to UNIX- or ZENIX-based operating systems are planned.

Cellular Kinetics Models of the Human Immune System (An Investigation of HIV-like Infections in a Model Immune System and its Response to Opportunistic Pathogens)

John E. Fletcher, Ph.D.
with J.J. Bailey, M.D., R.I. Shrager, M.S. (DCRT/LAS); H.C. Lane, M.D. (NIAID/DCD)

LAS is investigating models of the human immune system and the kinetics of its complex

interacting components (i.e., precursors, CD4+ T-cells, B-cells, T-cytotoxic cells, T-suppressor cells, natural killer cells, monocytes, interleukins, and lymphokines) by means of a system of nonlinearly coupled differential equations. An appropriately constructed and validated model should help in the design of experiments and interventions that can guide the use of treatments and vaccines and promote understanding of how the immune system might be manipulated to increase its effectiveness in preventing or neutralizing pathogenic infections.

In FY90, mathematical models of the interaction of the human immunodeficiency virus (HIV) with human T-cells *in vitro* were developed and their stability properties are being studied. The model represents combining healthy and HIV-infected CD4+ T-cells in a culture designed to keep the healthy cells growing. The model featured direct cell-to-cell transmission of HIV, transmission via the supernatant culture media, delay of infectivity, delay of viral synthesis, nutrient limitation, and quasi-steady state kinetics. Each of these features affect the model's stability characteristics.

Future Tools, Trends

The success of these studies will require close collaboration with laboratory scientists to obtain experimental validation. Experimental correlations with such models are currently being explored with NIAID investigators and their contractors.

Mathematical and Computational Methods for Solving Nonlinear Equations

Richard I. Shrager
with G. Weiss, Ph.D. (DCRT/LSM); J. Fletcher, Ph.D., J. Bailey, M.D. (DCRT/LAS); P. McPhie, Ph.D. (NIDDK/LBM); I. Levin, Ph.D. (NIDDK/LCP); R. Berger, Ph.D., R. Hendler, Ph.D. (NHLBI/LC); R. Winslow, M.D., K. Vandegriff, Ph.D. (LAIR Blood Res.); M. Gitterman, Ph.D. (Bar-Ilan University); M. Lewis, Ph.D. (NCRR/BEIP); G. Chnurny and G. McGregor (NCI/FCRDC); S. Gill and D. Ownby (University of Colorado)

This project provides NIH investigators with mathematical tools for insight, analysis, and solution of complex equations that arise in modeling biological systems. LAS has developed mathematical methods that are accessible to investigators from many disciplines and has released the resulting software packages to the research community as general research tools. Advice on the use of certain commercial mathematical software packages is also offered.

FY90 Progress: External Collaborative Projects

Conformational Changes in Hemoglobin Binding (with R. Winslow and K. Vandergriff, Letterman Army Institute of Research; S. Gill and D. Ownby, University of Colorado). Conformational changes in hemoglobin-oxygen binding was studied using singular value decomposition (SVD) and the most precise optical spectra that currently available equipment could provide. Once experimental procedures are refined, changes in molecular conformation due to oxygen saturation may be detected. Data are currently being analyzed to improve experimental design.

Thermal Unfolding of Swine Pepsinogen (with P. McPhie, NIDDK). A paper on the thermal unfolding of swine pepsinogen was completely rewritten to incorporate results from SVD and single wavelength analyses and a new multi-wavelength least squares analysis.

Infrared (IR) Spectra of the ATP-ADP-PO4 System (with R. Berger, NHLBI). At the 1-40 micromolar level, IR spectra of the ATP-ADP-PO4 system were analyzed by partial least squares,

but a repeatability problem developed. Initial calculations showed promise but exhibited about 25% error, in part because the small signal-to-background ratio placed heavy demands on the precision of the experiment. If this IR method can be refined, it will have many other uses, including non-invasive clinical applications. A poster session was given on this work at the Biophysics meeting, and the abstract was published.

Anti-Tumor Drug Spectroscopy (with G. Chmurny, G. McGregor, NCI/FCDRC contractors). Since certain anti-tumor drugs have intricate NMR spectra, one such model was programmed for use in a least squares analysis.

FY90 Progress: Internal Projects in DCRT

Periodic Fields (with G. Weiss, DCRT/PSL; M. Gitterman, Bar-Ilan University). A paper on the influence of noise on systems described by the Mathieu equation was written and published. This work has application in the design of experiments involving electric fields and diffusion. The results could have application to the design of electro-focusing experiments.

Modeling Systems. Efforts were made to gain experience with several small-computer systems, including IBM and Apple equipment, DOS and UNIX operating systems, and EDLIN, PE2, and WordPerfect word processors. Major effort went into developing MATLAB, MATHEMATICA, MLAB, and ML/e mathematical software, the last two under contract to NCI. The several beta test versions of the software are currently being field tested with selected investigators. LAS has participated by reporting bugs and suggesting numerous improvements in pre-release versions of the software. After the software is released, LAS and LSM staff will provide advice and consultation on the proper use of the software and interpretation of computer results.

Future Tools, Trends

In FY91, the redox states of *iron-in-cytochrome* oxidase will be monitored by Raman spectroscopy, and other aspects of the enzyme will be monitored by optical absorbance. The pooled data will be analyzed by SVD. In addition, a model of a *complex NMR spectra* will be used to model data from an anti-tumor drug experiment. The model may be further refined and/or generalized for other substances after it is validated. The *dissociation of ricin*, a highly toxic 2-subunit protein from the castor bean, will be studied by least squares applied to numerically transformed ultracentrifuge data. Further refinement of experiments involving *hemoglobin and ATP* is also expected. As the *PC-MLAB* and *ML/e* software packages evolve and become more reliable, LAS staff involvement will provide continued support, teaching and consultation.

Mathematical Models of Binding Equilibria

John E. Fletcher, Ph.D.
with E.W. Richards, Ph.D. (Baptist Research Centers, Dept. of Research, Birmingham, Alabama)

This project studied mathematical models of ligand-receptor and ligand-macromolecule binding studies at equilibrium. Appropriateness of various model-fitting criteria were studied and general guidelines and computational algorithms were designed for computer-aided interactive model fitting.

During FY90, a revision of the binding model concepts for 16-[9-anthroyloxey] palmitoyl-CoA to BSA was conducted. The manuscript was

given final revisions and was accepted for publication. No further activity is anticipated in this project.

Computer-Aided Analysis of Electrocardiography

James J. Bailey, M.D.
with M.R. Horton, M.Sc., E.W. Pottala, Ph.D. (DCRT/LAS); G. Campbell, Ph.D. (DCRT/LSM); D. Levy, M.D., J. Norman, Ph.D. (Framingham Heart Study, NHLBI); W.K. Haisty, Jr., M.D. (Cardiology Department, Bowman-Gray School of Medicine); M. M. Laks, M.D. (American Heart Association)

These studies are evaluating the prognostic power of the electrocardiogram when analyzed by advanced computer methodology and the predictive accuracy of diagnostic criteria when implemented in ECG computer programs. The use of well-documented populations and multivariate statistical techniques in designing new criteria are also under investigation. Studies have been pursued in collaboration with NHLBI and Framingham personnel and with the European consortium on Common Standards for Quantitative Electrocardiography (CSE).

The LAS and the DCRT Laboratory of Statistical and Mathematical Methodology (LSM) developed statistical methodology for determining the relative merits of diagnostic statements using McNemar's test for significance of differences in correlated results. This method was extended, and the well-documented CSE ECG data base was used to compare computer analysis by several different programs with visual interpretation by several expert readers. The data included both normal cases and cases representing five different categories of abnormality.

LAS and LSM developed methodology for comparing the relative diagnostic power of ECG algorithms as reflected in Receiver Operating Characteristic (ROC) curves. This methodology was demonstrated with data from the Framingham Heart Study and published. LAS and LSM further developed a method for constructing "fuzzy" ROCs in which cases with intermediate values for left ventricular mass were assigned a degree of membership to left ventricular hypertrophy. We have adapted a bootstrap method for estimating variance because the area under the fuzzy ROC curve was not amenable to a simple theoretical analysis as was the nonfuzzy ROC. A preliminary report was presented at the 15th Annual Conference of the International Society for Computerized Electrocardiology in April 1990.

Scientists in LAS also developed methods for digital signal processing in electrocardiology. Papers were published which describe the construction of bilinearly transformed, null-phase (BLT/NP) filters which maximize suppression of baseline wander with minimal displacement of the ST segment. LAS staff constructed a bilinearly transformed, null-phase, notch filter to suppress powerline interference; this work was presented at the 15th Annual Conference of the International Society for Computerized Electrocardiology.

Future Tools, Trends

In the coming fiscal year, the method of using fuzzy ROCs will be again applied to data from the Framingham project to investigate whether adjusting ECG voltages for age dependence can improve an LVH diagnostic algorithm. The BLT/NP filter method will be extended and further applied to suppress high frequency noise.

Analysis of Physiological Signals

E.W. Pottala, Ph.D.
with J.J. Bailey (DCRT/LAS); J.A. Dvorak (NIAID/LPD); K. Rasmussen (NICHD/LCE)

This project involves developing and applying microcomputer-based signal processing techniques to analyze physiological signals, e.g., electrocardiogram, electromyogram, and electroencephalogram. The LAS microcomputer-based systems provided a general-purpose analog-to-digital conversion facility and an ability to pre- or postfilter the signals with a variety of analog and digital techniques.

Investigators at the Laboratory of Comparative Ethology, NICHD, are collaborating in a study of heart rate variability (HRV) in free-living simians. In FY90, methods were improved to deal with muscle noise, artifact, extrasystole, and other confounding factors that may interfere with obtaining an accurate measure of HRV in the frequency domain. The analog-to-digital conversion and all analysis of the data have been implemented on a PC-XT using Microsoft QuickBasic-4.

The development of new filtering techniques continued on the IBM/XT and the Macintosh II using the MATLAB mathematical modeling software package. New filters for preprocessing ECG data to suppress powerline interference without distortion of the ST segment were designed and tested against other commonly used methods. Reports on these studies were presented at the 15th annual conference on "Computerized Interpretation of the Electrocardiogram."

MATLAB on a microcomputer permits a remarkable facility for interactive design of digital processing techniques and display of results. LAS is developing a single-pass filter to suppress baseline wander which will not seriously alter the ST segment and will operate in near real time (i.e., delayed about 0.16 second). This will be important for ECG monitors in intensive care units, where the filters currently in use to suppress baseline wander seriously distort the ST segment and can lead to false diagnoses of myocardial ischemia or injury.

The signal processing hardware completed its transition from the obsolete LSI-11 minicomputer system to new Apple MacIntosh equipment. The analogy-to-digital and processing functions were improved with the new Spectral Innovations signal processing hardware.

Future Tools, Trends

Most signal processing activities in FY91 will concentrate on the Macintosh IIx, a state-of-the-art system with excellent display and interactive capabilities that can handle large data arrays. With the signal processing board, it can performa Fast Fourier Transform on 1024 data points in 3.4 msec, as well as high-speed analog-to-digital conversion and additional signal processing as necessary. Single ECG filters will be developed on the newer Macintosh system, and special Macintosh applications programs will be developed in the C language to generalize and simplify the use of the array processing board.

Computer Systems for Nuclear Medicine

Margaret Douglas, B.A.
with J.J. Bailey, M.D. (DCRT/LAS); S.L. Bacharach, Ph.D., M.V., Green (CC/Nuclear Medicine); R.O. Bonow, M.D. (NHLBI/CB)

LAS developed systems for computer-based mathematical analysis, pattern recognition and image processing in support of diagnostic activities in the Nuclear Medicine Department of the Clinical Center and collaborating Insti-

tutes. Applications include: estimation of parameters of ventricular function extracted from radionuclide ventriculography; tumor detection and imaging utilizing radio-labeled monoclonal antibodies; functional imaging of the central nervous system using PET scan technology; and the correlation of function with structure by superposition of PET, CT and MRI scans in three dimensions.

FY90 Progress

Over the past four years, LAS, in collaboration with the Nuclear Medicine Department of the Clinical Center, has designed and specified a general-purpose image processing system called MIRAGE. Programming was performed by contractors supervised by LAS. The completed basic system has been ported to several other NIH computer systems, and a three-dimensional imaging module for MIRAGE has been completed.

LAS evaluated various general-purpose image processing facilities for the Nuclear Medicine Department, including a Sun 4 Workstation with TAAC-1 hardware and software and the Mayo Clinic image processing package ANALYZE. Modifications and additions were designed for the TAAC-1 system to make it more responsive to the needs of the Nuclear Medicine Department, and both the TAAC-1 and ANALYZE systems were used to investigate problems of advanced three-dimensional visualization of medical images (eg. CAT, MRI, and PET scan).

Other LAS activities in applied image processing included: work with a task force to generate a software, hardware, and user interface specifications for a general purpose image processing system for use in a variety of environments at NIH; investigation of the use of neural networks in the problem of image segmentation and classification; campus-wide consultation and collaboration on image processing, especially in PC-based image processing; teaching PC-based classes in image processing and developing a review of the current state of PC-based image processing software; serving as image processing representative to groups involved with NIH campus computer connectivity; and investigation of PC Computer Aided Design. LAS personnel also sponsored seminars through the NIH-wide Image Processing Group and collaborated in publishing the Directory of NIH Image Processing Facilities.

Future Tools, Trends

As the Nuclear Medicine Department has several systems which collect three-dimensional image data, LAS will continue to investigate three-dimensional visualization, multimodality registration, and development of fast interactive algorithms for analysis of large volumes of data. LAS will participate in the design of a Clinical Center image transmission network; design and specify standards for image processing workstations; help to develop a consensus approach to medical image archiving issues; and contribute to the design and functional specifications for image processing software for a system designed to be the standard system for macroscopic image processing (X ray, PET, MRI, CAT scans, etc.) at NIH.

Computer-based Analysis and Image Processing in Electron/Light Microscopy and X-ray and Electron Energy Spectroscopy

Margaret Douglas, B.A.

This project was directed toward developing computer-based mathematical and statistical analyses for pattern recognition and image processing of data, principally for x-ray microg-

raphy, electron energy loss spectra (EELS), and the electron/light microscopy of biological specimens.

A PC-based system was developed for investigation of micrographic images and used to analyze micrographs of rat spinal cord to determine parameters of injury. The system served as a test bed for NIH researchers wishing to learn the feasibility of a PC-based image analysis system.

Other applied image processing activities included investigating automated cell classification from micrographs and reviewing the field of medical image processing of micrographs for an invited journal review article.

A Macintosh-based system was acquired and shared to investigate problems in both image and signal processing. Several image processing and graphics packages for both the PC and Macintosh systems were evaluated.

Future Tools, Trends

Future efforts will include increased use of the Macintosh for image processing, investigation of the feasibility of the Macintosh as an image processing workstation, and purchase of a faster Macintosh dedicated for the use of image processing. The UNIX operating systems for the PC (AIX) and Macintosh (AUX) will be evaluated.

Interface for Computer Analysis of Molecular Databases

Martha R. Horton, M.Sc.
with E. Tyler, B.Sc., J.J. Bailey, M.D. (DCRT/LAS); P.R. Krause (NIAID/IRP/LCI)

This project was developed to design and implement a much-needed interface between researchers using molecular biology data bases and the computer programs that analyze these data. To use existing data bases and programs, researchers must know the calling requirements of the various programs and the operating formats of the data bases. This project is directed toward constructing a computer software interface to enable actual manipulation of the available data with a minimum of inconvenience to the researcher.

In FY90, an extensive review of computer algorithms to compare molecular (polynucleotide or polypeptide) sequences and/or subsequences resulting from restriction enzyme analyses was completed. A manuscript describing this work is in press.

Future Tools, Trends

Further work in this area will be performed primarily by the newly formed computational molecular biology unit.

Publications and Presentations

Bailey, J.J.; Berson, A.S.; Garson, A. Jr.; Horan, L.G.; Macfarlane, P.W.; Mortara, D.W.; Zywietz, C.: Recommendations for standardization and specifications in automated electrocardiography: bandwidth and digital signal processing. *Circulation* 81, 730-739 1990.

Boverman, J.F.; Burgess, R.C.; Bourgeois, B.F.D.; Turnbull, J.P.; Horton, M.R.: On-line anticonvulsant pharmacokinetics analysis on a multi-user system. *Proceedings of the 11th Annual International Conference of the IEEE Engineering in Medicine and Biology Society,* September, 1989.

Burgess, R.C.; Braun, D.S.; Boverman, J.; Edwards, C.R.; Horton, M.R.: A flexible signal processing interface for a neurophysiologic ICU monitoring system. Accepted for presentation, 5th European Congress of Clinical Neurophysiology, September, 1990.

Campbell, G.; Douglas, M.A.; Bailey, J.J.: Nonparametric comparison of two tests of cardiac function on the same patient population using entire ROC curve. In: Ripley K.L and Murray A. (eds.) *Computers in Cardiology,* No. 88CH2733-4. IEEE Computer Society, Washington, D.C 267-270 1989.

Campbell, G.; Levy, D.; Bailey, J.J.: Bootstrap comparison of fuzzy ROC curves for ECG-LVH algorithms using data from the Framingham Heart Study. *J. of Electrocardiology* 1990 (in press).

Campbell, G.; Levy, D.; Lausier, A.; Horton, M.R.; Bailey, J.J.: Nonparametric comparison of entire ROC curves for computerized ECG left ventricular hypertrophy algorithms using data from the Framingham Heart Study. *J. of Electrocardiology* 22 (supplement), 152-157 1989.

Douglas, M.A.; Trus, B.L.: An introduction to image processing in medical microscopy. *Medical Progress Through Technology* 15, 109-140 1989.

Fletcher, J.E.: Introduction to software for numerically solving reaction-diffusion-transport models. DCRT Seminar Series, April 1990 (LAS manual with software distribution).

Gitterman, M.; Shrager, R.I.; Weiss, G.H.: Influence of noise on systems described by the Mathieu equation. *Physics Letters A.* 142, 84-87 December 1989.

Gradwohl, J.R.; Pottala, E.W.; Horton, M.R.; Bailey, J.J.: Comparison of two methods for removing baseline wander in the electrocardiogram. *Computers in Cardiology,* No.88CH2733-4. IEEE Computer Society Press, Washington, DC. 493-496 1989.

Mermelstein, A.L.; Carducci, J.S.; Shrager, R.I.; Berger, R.L.: Detection of ATP, ADP, and Phosphate in solution by near infra-red Spectroscopy. *Abstracts in Biophysics* 57(2), 54a 1990.

Pottala, E.W.; Bailey, J.J.; Horton, M.R.; Gradwohl, J.R.: Suppression of baseline wander in the ECG using a bilinearly transformed, null phase filter. *J. of Electrocardiology* 22 (supplement), 243-247 1989.

Pottala, E.W.; Horton, M.R.; Bailey, J.J.: Suppression of powerline interference in the ECG signal using a bilinearly transformed, null phase, notch filter. *J. of Electrocardiology* 1990.

Richards, E.W.; Hamm, M.W.; Fletcher, J.E.; Otto, D.A.: The binding of Palmitoyl-CoA to bovine serum albumin. *Biochemica et Biophysica Acta* 1044, 361-367 1990.

Shrager, R.I.: Improvement of SVD properties by empirical rank-1 weights. Invited lecture at FACSS 16th Annual Meeting, Chicago, Illinois, October 2, 1989.

Tyler, E.; Horton, M.R.; Krause, P.: A review of algorithms for molecular sequence comparison. *Computers in Biomedical Research* 1990 (in press).

Willems, J.L.; Campbell, G.; Bailey, J.J.: Progress on the CSE diagnostic study—applications of McNemar's test revisited. *J. Electrocardiology* 22 (supplement), 135-140 1989.

LSM

Laboratory of Statistical and Mathematical Methodology

James E. Mosimann, Ph.D., Chief

The Laboratory of Statistical and Mathematical Methodology (LSM) combines research in mathematical statistics, mathematics, and computer and information science with collaboration and service in these areas for NIH researchers and administrators. LSM staff interact with all NIH Institutes, other Federal agencies outside DHHS, and biomedical researchers worldwide.

LSM has 13 full-time professionals, including research mathematicians, mathematical statisticians, computer systems analysts, and programmers. There are four sections:

Statistical Software Section (SSS) provides consultation to and collaboration with NIH researchers and administrators in all computational aspects of biomedical data analysis, including selection and support of large systems and packages.

Statistical Methodology Section (SMS) works closely with the Statistical Software Section to provide biostatistical consultation, DNA analysis, and computer graphics software tools. Another major function of the SMS is to perform independent research.

Biomathematics and Computer Science Section (BCS) performs independent research, provides consultation in mathematics and on software used for scientific printing, and develops LSM computer network facilities.

Medical Information Science Section (MIS) investigates and develops methods for applying information and computer science to medical language data processing.

Computation in FY90

LSM activity offers statistical, mathematical, and other scientific systems and packages to the NIH user community and evaluates new systems and packages for suitability to NIH needs. Computer systems and packages supported by LSM are shown in Table 1. Use of these statistical packages at NIH since 1975 is shown in Figure 1. Average monthly use remained at a high level in FY90.

As in previous years, the SAS statistical and data management system was extensively used at NIH, with an average of 79,600 accesses per month via the IBM System 370. SPSS-X was accessed around 4,100 per month, and the BMDP package was accessed an average of 550 times per month.

LSM support included maintenance of the system or package and provision of adequate documentation, including NIH computer system changes, system or package updates, and corrections. It also included rapid response to queries concerning user access to the most used systems and packages. The SSS staff answered over 6,500 calls for SAS assistance, including requests for information on job control language, program parameters, and other operating system procedures, as well as assistance in interpreting results.

Recognizing the importance of teaching effective use of systems and packages to biomedical researchers and other NIH users, LSM maintained a substantial program of short courses, documentation preparation, and informational talks and articles. LSM taught eight introductory courses on SAS, three on SAS/GRAPH, and three on Permanent SAS Files. Seminars for DNAdraw were also offered, as well as three statistical courses on Recurrent Problems in Data Analysis.

Other software supported by LSM had more limited use. Usage ranged from an estimated

Table 1. Systems and Packages Supported by LSM

SAS, SAS/GRAPH, SAS/ETS, SAS/OR, SAS/FSP, SAS/AF, SAS/IML, SAS/CBT
Vendor: SAS Institute, Inc. A batch and interactive IBM S/370 system for statistical analysis, with extensive file manipulation capabilities and graphics, also in interactive mode on the IBM PS/2 and compatibles.

SPSS-X, SCSS, SPSS-PC
Vendor: SPSS, Inc. A system for univariate and multivariate statistical analysis with file handling capabilities, in batch mode on the IBM S/370, and interactive mode on IBM S/370 and PS/2 and compatibles.

BMDP
Vendor: BMDP Statistical Software, Inc. A collection of IBM S/370 batch programs for univariate and multivariate statistical analysis.

IMSL (International Mathematical and Statistical Libraries)
Vendor: IMSL, Inc. An extensive collection of FORTRAN routines for statistical and mathematical analysis, for Convex and IBM S/370 batch computation.

MSTAT I
Source: DCRT staff. IBM S/370 batch programs and subroutines for mathematical and statistical analysis.

GLIM (Generalized Linear Interactive Modeling)
Vendor: Numerical Algorithms Group, Inc. An IBM S/370 batch and interactive system for analysis of linear statistical models.

CART (Classification and Regression Trees)
Vendor: California Statistical Software, Inc. An IBM S/370 batch and interactive system for tree-structured regression and classification analysis.

DNAdraw
Source: LSM staff. A menu-driven PC program for preparing publishable displays of DNA sequences.

VMAP
Source: LSM staff. An IBM S/370 batch program used for printing scientific text and diagrams. It is used in conjunction with the WYLBUR text editor and IBM 6670 laser printers.

1,000 MLAB sessions, 400 GRAPH sessions, and 200 VMAP accesses per month to relatively few sessions for such specialized programs as REDUCE, DNALAB, GLIM, and CART. A new version of DNAdraw for IBM-PC compatible microcomputers that formats, annotates, and displays DNA sequences for publication was distributed in FY90.

Consultation and Collaboration in FY90

LSM provided consultation in a wide range of scientific fields and in interpreting the results of statistical and other scientific computation. Staff members also provided consultation in statistics, biomathematics, and computer analysis of DNA sequences. Consultations were brief or extensive, depending on the complexity

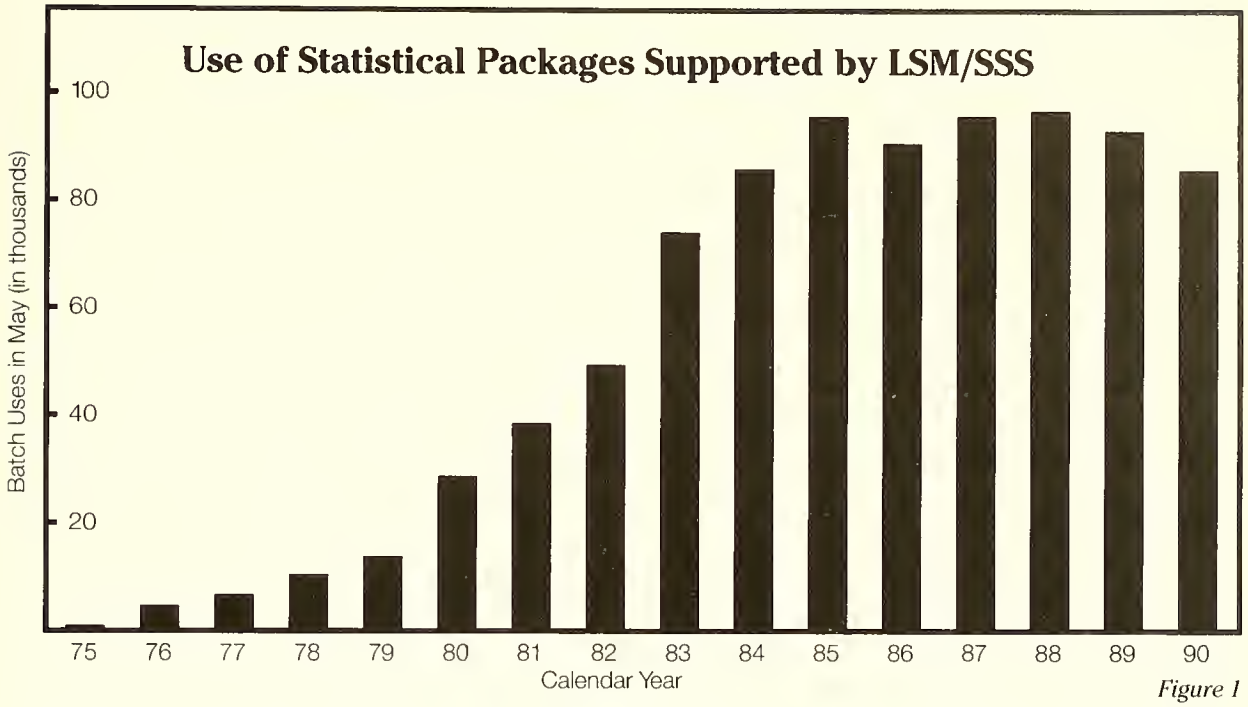


Figure 1

of the statistical or scientific research; some evolved into long-term collaborations resulting in scientific publications.

Consultations on the use of the DNAdraw program for producing publication quality drawings of DNA sequences included Dr. S. Cheng (NCI/DCBD), Dr. G. Einhorn (NIAID/BRB), Dr. W. Leonard (NICHD/CBMB), Dr. J. Adams (NIAID/LPD) and Dr. J. Mill (NINDS/LMB).

Collaborative statistical efforts included the following:

- Dr. R. Callahan, Dr. C. Cropp (NCI/LTI). For a study of loss of heterozygosity in various regions of the human genome in the tumor tissue of women with breast cancer, LSM provide a novel methodology to detect a low level of dependence among oncogene sites on different chromosomes. A manuscript has been accepted for publication by the Proceedings of the National Academy of Sciences. A second study concerned the incidence of the oncogenes *int-1* and *int-2* in two strains of mice and in a third strain fed breast milk by mothers of either of the first two. Proportions of *int-1* and of *int-2* were compared between each original strain and the third strain. LSM also assisted in the study the dependence of *int-1* and *int-2*.

- Dr. K. Bayne (NCRR/OD). In a study of the behavior of caged rhesus monkeys (*Macaca mulatta*), LSM provided a novel nonparametric analysis to evaluate the effectiveness of introducing into the cage a device to reduce abnormal behaviors and to foster grooming and foraging. A paper has been accepted and another is in preparation.

- Dr. D. Rabin (NICHD/DEB). In a study of normal women and women with Premenstrual

Syndrome (PMS), LSM provided statistical advice, including stepdown multivariate procedures, for analyzing an experiment that studied the response of adrenocorticotropic hormone (ACTH) and cortisol to ovine corticotropin releasing hormone (CRH) in both luteal and follicular phases.

- Dr. P. Carroll (NIDDK/LCBG). For a study of the effect of doses of the immuno-suppressant drug FK506 and of cyclosporine A on insulin production using the animal islets, LSM provided a nonparametric analysis using multiple comparison procedures to compare the different drugs.

- Dr. I. Feuerstein (CC/DR). In a study of the tracheal and pulmonary mean densities in normal patients, LSM provided regression and correlation methodology to study the effect of patient age and adjust for patient variability. This methodology will be applied to patients with emphysema.

- Dr. M. Bolander, Dr. M. Joyce (NIAMS/ARB). A study of the effect of exogenous transforming growth factors beta-1 and beta-2 on bone formation and fracture healing was undertaken using the rat model. LSM provided an unbalanced three-way analysis of variance model studying the effect of growth factor, dose and day of sacrifice on induced mass of intramembranous bone, endochondral bone, and cartilage, using a log transformation to stabilize the variance of the dependent variable. A second study examined the effect of estrogen on fracture healing and bone formation in the femurs of rats, incorporating dose of estrogen, estrogen level in the blood at sacrifice, and the initial and final weights under a cohort feeding model.

- Dr. S. Marx (NIDDK/MD). In a study of patients with hyperthyroid hyperplasia, LSM provided an analysis to evaluate the size and shape of the four hyperthyroid glands based on approximated dimensions, shape and weight. The study included estimation of volume on clay models by two observers to quantify sources of variability; the insight gained here was applied to glands in patients at time of surgery.

- M. Flack, G. Cutler (NICHD/DEB). In a study of the differential diagnosis of Cushing's syndrome, based on two tests of hormonal dysfunction, Cochran's Q statistic and receiver operator curves (ROC) were applied to determine the superior test procedure, in terms of percent correct diagnosis, specificity, and sensitivity. The resulting paper was submitted to the New England Journal of Medicine.

- S. Yanovski, J. Yanovski (NICHD/DEB). In an intensive study of the telephone (on-call) decision-making process for primary care family practice and pediatricians, three scenarios were presented to 142 physicians in a pre-arranged "on-call" setting. Responses were assessed in terms of final correct diagnosis, number of critical areas addressed, and time to decision. The respondents were grouped in several ways (level of training, academic or private practice, years of practice, etc.), and the responses were examined using contingency tables, parametric and non-parametric analysis of variance, and Goodman and Kruskal's gamma statistic as a measure of regression on ordinal category data. A paper was accepted for presentation at the annual meeting of the American Academy of Family Physicians.

• G. Cutler, K. Barnes (NICHD/DEB). Normal and Turner's syndrome children were studied using two measures of response over time to growth hormone administration. The children were also grouped by age, less than or equal to or greater than nine years old. Multivariate repeated measures analysis was applied for these unbalanced data, and two statistics were used to evaluate the relative potency of the dose responses. First, an adaptation of an existing procedure was used to assess relative potency; then an original statistic was devised to take into account the correlation (repeated measurements) structure of the data. A paper is in preparation.

• N. Martin, R. Rubinfeld (George Washington University/Department of Ophthalmology). The link was examined between two eye syndromes: giant papillary conjunctivitis and meibomian gland dysfunction. In practice, it has been found that one or the other gets properly treated but not both, and the study demonstrated the strong link between the two, concluding that both required concurrent treatment for most effective management. Goodman and Kruskal's measure of ordinal regression was applied to assess the strength of the relationship. An abstract was presented at the annual meeting of the American Academy of Ophthalmology, and a paper is in preparation.

• E. Deranzo (CC/Bioethics). In a study of the apparent increase in age-related discriminatory language in the popular literature, specifically in skin care product advertising for the years 1969 through 1988, it was found after transforming the data to adjust for the relative increase in advertising pages generally over these years, that a quadratic equation fit the age-biased language occurrence rate well. This suggests that while a decrease in age related discriminatory language occurred in the early 1970's, such language and attitudes are significantly on the increase in recent years. An abstract was presented at the Seventh National Forum on Research in Aging, and a paper is in preparation.

• D. Johnson (CC/Bioethics). A lengthy questionnaire survey study was conducted among more than 400 clinic physicians from several of the institutes at NIH to examine the differing degrees of doctor-patient treatment and progress disclosure rates. Also examined were the situational, context-sensitive distinctions made by the physicians and their rationales for the amount of information disclosed, when, and to whom. A paper is in preparation.

• J. Sorkin (NIA/IRP). Aspects of the Baltimore Longitudinal Study on Aging were studied, particularly the question of how to assess relative changes in height for an individual over a period of 10 and 14 years, as compared with height changes for the patient's cohort. A new technique, using the covariant derivative of differential geometry, was devised, with an appropriate statistic, to measure the height velocity change of an individual compared to the average velocity change of a group. Data preparation and programming of the new statistic continues.

• J. Tal (NICHD/LCE). A study was made of the frequency and kind of behavior change of a group of mothers in an interactive learning setting for children in two age groups, six months and one year old. As only the changes to a new behavior were considered (not the time between changes) and as only lagged data was collected (so that an individual mother's

complete record of changes and behaviors was lost to analysis), a new statistical technique was required. The method devised used a first- and second-order Markov process and maximum likelihood to estimate the transition matrices. Programming continues, and a paper describing the new method is in preparation.

- Dr. B. Jones (NIMH/LPP). LSM provided expert statistical advice on the selection of cluster analysis algorithms in the study of MMPI profiles of mental health problems.

- Dr. C. Venter, Dr. A. Kerlavage (NINDS/LMCN). In collaboration with J. Powell (CSL), LSM evaluated current computer needs for gene sequencing work and provided expert advice for future planning. In addition, the feasibility of applying laboratory notebook software for database needs was studied.

- Dr. A. Cheever (NIAID/LPD). LSM provided statistical assistance in a study of egg persistence in tissue in schistosome infections after treatment. A paper is in preparation.

Research and Development in FY90

Research projects in LSM varied widely; they included statistical methodologies for biomedical applications, scientific computer printing and graphics, research in mathematical and statistical methods, and language processing for medical information systems.

During FY90, SMS staff members were active in mathematical statistical research in a variety of areas.

A comprehensive study of Receiver Operating Characteristic (ROC) curves that began in FY88 was continued. The ROC curve is an important tool in the evaluation of diagnostic biomedical tests. Development of ROC analysis with fuzzy data was continued, using the highly computational bootstrap methodology for statistical inference. Parametric models using Lomax and Pareto distributions were implemented with maximum likelihood computer programs. Presentations on these topics were made at several national and international meetings. Investigations in nonparametric statistics continued, and relationships of standard statistics with the ROC plot were explained. Work has begun on procedures for nonparametric discrimination and classification.

Studies of size and shape variables continued to provide methods for studying random proportions or ratios of common occurrence in biomedical data. Studies of patterned covariance matrices continued. One paper on transforming ratios and proportions to obtain approximately normal distributions that permit statistical analysis is in press (with M. V. Ratnaparkhi). The methods were applied in a second paper (with C. V. Wiseman, J. J. Gray, A. H. Ahrens) on cultural expectations of thinness in women in relation to eating disorders.

Mathematical methods were utilized to investigate problems of statistical inference in the presence of quantum uncertainty. An invited paper on this topic is in preparation for a review journal. Applications of this new statistical methodology arise in many areas of modern biomedical technology, including positron emission tomography scan and bioluminescence assay techniques. A monograph on algebraic methods in statistics has been completed. The study of the problem of optimal unbiased estimation of variance components in mixed models with possibly unbalanced data

was continued. A number of invited lectures on these topics were given in the past year.

During FY90, the staff of the Biomathematics and Computer Science Section spent considerable time upgrading the local area network facilities available to LSM and LAS staff. The current configuration includes six PS/2 Model 80s, four PS/2 Model 60s, three Macintosh IIs, and four 3Com3Station microcomputers connected to two LSM/LAS network servers, and one Sun 3/50 workstation connected to the Advanced Laboratory Workstation local area network. Another server, several Macintosh II microcomputers, and many new or upgraded network peripherals are planned. A contract for testing commercial software for mathematics is in progress.

Biomathematics and Computer Science Section research included development of a methodology for relations in algebraic structures and for new systems of corresponding dual relations. Many connections with known theory of universal algebra were obtained. Applications to computer methodology and discrete mathematics as applied to biomedical research are under study. A research paper was presented at a symposium on algebra, lattices, and logic in Laugervatn, Iceland.

In the Medical Information Science Section, research in medical linguistics continued at the same level. A parsing methodology providing direct representation of semantics in a predicate calculus formalism is being implemented. A separate project for development of a comprehensive lexicographic data base will be continued and used in collaborative lexicographic research. LSM continued collaboration with the Laboratory of Pathology of NCI/DCBCD and with DCRT/DMB, using LSM's automatic encoding system to maintain the surgical pathology report data base of the Clinical Center.

Future Plans: Software Development, Contracts, Network Development

LSM's high level of support for IBM S/370 statistical software systems will be continued, as will support of GRAPH, DNAdraw, and VMAP for use on the PCB network.

LSM will continue its contribution to software testing and development for two DCRT local area computer networks: a network supporting IBM-compatible and Macintosh microcomputers, and the advanced laboratory workstation project of CSL supporting Unix-based workstations. Support of SAS in a PC/networking environment will also continue.

Research Projects

Automated Data Processing of Medical Language

G. Dunham; A.W. Pratt; M.G. Pacak (Guest Researcher); S. Harper (LSM)
with E. Jaffe (NCI/DCBD/LP); D.E. Henson (NCI/DCPC); A. Moshell (NIAMS/EP/SDB)

The major objective of the project was developing methods for automatically processing natural medical language. Research in medical linguistics included the development of paraphrasing rules for medical noun phrases, a semantic grammar, and a generalized morphological analysis approach suitable for general language and data compression as well as medical language morphology. Development continued on the Lexicographic Environment Software, a platform for medical lexicography and the development of medical language processing.

Research focused on theoretical and developmental issues of integrating medical text and the existing theorem prover for relevance/entailment predicate calculus with the Lexicographic Environment Software. Examples included optimal text compression consistent with projected learning algorithms and the exploitation of common English question syntax in learning semantics and syntax. All components of the lexicographic software system were designed to facilitate the adjoining of learning procedures to bear some of the task of developing detailed medical dictionaries and language processing procedures.

Collaboration continued on the Clinical Information Utility with the Laboratory of Pathology (NCI/DCBD/LP) and with DCRT/DMB to use LSM's automatic encoding system to maintain and improve the data base of Clinical Center surgical pathology reports. Progress was made on obtaining and automatically encoding NIH surgical pathology data from the Clinical Center Medical Information System (MIS) files.

A project was begun to explore the application of collocation and display techniques, in collaboration with NCI/DCPC and NIAMS/EP/SDB. Originally designed for linguistic research, the project looks toward the achievement of consensus by experts on the terminology and criteria for the diagnosis of borderline malignant melanoma.

Computer Graphics and Applications
M. B. Shapiro

The main objective of this project was the application of computer graphics and related methods to NIH research problems.

The DNAdraw program for the IBM-PC was completed, tested, and distributed, with documentation, to a large number of potential users. With DNAdraw and an HP LaserJet printer, a researcher can easily prepare formatted, annotated, and highlighted sequences for publication, usually in less than an hour. Some work continues on the program for minor improvements and correction of errors.

In collaboration with K. Li, CC, a small software package for printing labels for medical records was developed.

Discrete Mathematics and Applications
G. Hutchinson, Ph.D.

The project objective was to develop mathematical theory and computational techniques using discrete mathematics (algebra, combinatorics, and graph theory), and to apply such methods to appropriate problems of biomedical research and computer science.

A method for defining systems of relations was generalized to categories possessing an image factorization system. Using category duality, systems of elements which are category duals of relations (coproduct congruences) were examined. It was determined that a variety of algebras support this dual construction if and only if the variety has the amalgamation property and the congruence extension property (both well studied). The congruence permutability property for varieties of algebras was shown to be equivalent to a regularity formula for the systems of relations, and to a category property. The result that congruence permutability implies congruence modularity was refined into several interrelated category results. Many new properties of dual relation systems were immediately obtained by dualizing proofs. A research paper was pre-

sented at an symposium on algebra, lattices and logic in Laugervatn, Iceland.

Work continued on the characterization of rings which possess the same associated lattice theory of modules.

Software development of the Scientific Printing Utility software based on previous research was continued by BCS section personnel.

Analysis of lattices and categories of modules will continue. The new initiative on systems of dual relations will continue, and applications to computer software technology and discrete mathematics as applied to biomedical research are being sought.

Multivariate Statistical Analysis

J. E. Mosimann, Ph.D.
with G. Campbell (DCRT/LSM); M.V. Ratnaparkhi, (Wright State University, Dayton, Ohio); Paul Pirlot (Universite de Montreal); E. Depiereux, (Facultes Notre Dame de la Paix, Namur, Belgium); Claire Wiseman (The American University, Washington, D.C.)

The objective of this project was the study of multivariate ratios and proportions. Studies were continued on random proportions and ratios, including those that follow a multivariate lognormal distribution as well as those following the distribution of a mixture of Dirichlet distributions. Patterned matrices that arise frequently with biological data, "persymmetric" (including Toeplitz) matrices were studied. In addition, a paper with M. V. Ratnaparkhi on normalizing transformations of ratios and proportions was completed. A second paper studying thinness in women in relation to eating disorders was completed (with C. Wiseman). These research results are being assembled for a research monograph on the subject.

Algebraic Methods in Statistics

J. D. Malley, Ph.D.

This continuing project has developed new methods in statistics that have a substantial reliance on algebraic methods and immediate applications for biomedical research. Recently obtained results were compiled in a research monograph that has been accepted for publication by Springer-Verlag in the "Lecture Notes in Statistics" series. Some of the problems treated in the monograph include: simpler, unified methods for mixed model analysis of variance; new and complete generalizations of the standard tests for statistical independence of sums of squares, given an arbitrary family of allowable, unknown covariance matrices; a solution to several forms of mixed model analysis given missing data, using an algebraic solution to the convergence problem for the well-known Expectation Maximization (EM) algorithm (the solution allows for arbitrary patterns of missing data, as well as for arbitrary forms of the unknown covariance matrix). Using the new methods, convergence is guaranteed for many of the most common missing data problems, and applications can be made to longitudinal and growth model problems, as well as to certain genetics studies and time-series problems.

Quantum Statistical Inference

J. D. Malley, Ph.D.
with J. Hornstein, Ph.D. (Naval Research Laboratory)

This interdisciplinary project investigated the uses of statistical inference and estimation for the data that obey the rules of quantum mechanics. Many rapidly developing areas of biomedicine and biotechnology are, in fact, either based on a special form of quantum

theory, called quantum optics, or could appreciably benefit from a quantum theory treatment. Some of these include: positron emission tomography scans; real-time, laser-based confocal microscopy of living cells and tissue; bioluminescent molecular tagging; and enhanced chemoluminescence. Other applications include optical communications, fiber-optic sensors, remote sensoring, and many others in applied physics, chemistry and biology.

The statistical, mathematical and quantum theory machinery needed was fairly intensive but did permit a nearly seamless transposition of many of the classical statistical methods into these new areas. The new methodology required is being prepared for publication as an invited article for Statistical Science.

Statistical Inference and Neural Networks
J. D. Malley, Ph.D.

This project investigated the possibilities of using the methods of computational neural networks, simulated annealing, and the Boltzmann learning machine for significantly improved computational power for statistical inference and estimation. A recently obtained result, for example, proved that the well-known EM algorithm is in fact equivalent to the operation of that form of a neural network known as a Boltzmann learning machine. More generally, the established method of maximum likelihood estimation is now known to be realizable as a simulated annealing procedure.

This new connection of classical statistical methods and massively parallel computation presages both greatly improved estimation, for example with difficult missing value problems, and also a deeper understanding of statistical estimation in general.

The new results are being readied for publication as an invited article for Statistical Science. At the invitation of Prof. D. Rubin, Chairman, Department of Statistics, Harvard University, the material may appear as a volume in the Chapman and Hall (Publishers) series of statistical monographs.

Nonparametric Statistics
G. Campbell, Ph.D.

A comprehensive study of the theory and application to biomedical research of Receiver Operating Characteristic (ROC) curves continued, with emphasis on nonparametric techniques for estimation and testing. This methodology can be used to compare two diagnostic tests via ROC areas when the data are ordinal categories or continuous variables. Also, the theory of ROC curves was extended to fuzzy data; instead of knowing that a patient is diseased or nondiseased, it is sometimes useful to view the patient's disease state to be a continuous variable between zero and one, where zero indicates normal and one indicates diseased. With such data, one can define weighted estimates of sensitivity and specificity and hence the fuzzy ROC curve. The area under the ROC can be calculated easily. This avoids the issue of bias, which is usually introduced when the two groups are not well separated and individuals without clear identification are dropped from the analysis. Work is being extended to the statistical properties of analyses for comparing areas under two fuzzy ROC's, as well as the important problem of restriction to specific sections of the ROC curve (for example, limiting to specificities greater than 0.5). The computationally intensive statistical methodology of the bootstrap was studied and applied to the

evaluation of two fuzzy ROC curves as well as to restricted sections. This powerful methodology depends on the repeated sampling with replacement from the observed data. A study of the Lomax and Pareto distributions with application to the problem of parametric ROC modeling of data from the same patients of diagnostic tests has been developed jointly with Dr. M. Ratnaparkhi. Maximum likelihood estimation of the parameters has been achieved by iteration using a computer program written expressly for the problem. Relationships between the empirical ROC curve and nonparametric statistics has been carefully explored, and work has begun on nonparametric discrimination and classification procedures.

Publications and Presentations

Ali, I.U.; Campbell, G.; Merlo, G.R.; Smith, G.H.; Callahan, R.; Lidereau, R.: Multiple genetic alterations in human breast cancer and their possible prognostic significance. In: *Cancer Cells*, Vol. 7: Molecular Diagnostics of Human Cancer, Cold Spring Harbor, New York, 399-403 1989.

Bayne, K.; Mainzer, H.; Dexter, S.; Campbell, G.; Yamada, F.; Suomi, S.: The reduction of abnormal behaviors in individually housed Rhesus monkeys (Macaca mulatta) with a foraging/grooming board. *American Journal of Primatology*, 1990 (in press).

Callahan, R.; Campbell, G.: Mutations in human breast cancer: an overview. In: *J. of the National Cancer Institute* 81, 1780-1786 1989.

Campbell, G.: An application of Lomax distributions in Receiver Operating Characteristic (ROC) curve analysis. At Annual Meeting of the Institute of Mathematical Statistics, Washington, DC, August, 1989.

Campbell, G.: Bootstrap comparison of fuzzy ROC curves for ECG-LVH algorithms using data from the Framingham Heart Study. At International Society for Computerized Electrocardiology, Virginia Beach, Virginia, April, 1990.

Campbell, G.: Mathematical and computer-intensive statistics in a biomedical environment. At Maryland Chapter of the American Statistical Association, Baltimore, Maryland, October, 1989.

Campbell, G.: Medical decision-making and the ROC curve: a nonparametric perspective. At Washington Statistical Society, January, 1990.

Campbell, G.; DeLeo, J.M.: Fundamentals of fuzzy Receiver Operating Characteristic (ROC) functions. In: Berk, K.; Malone, L.(eds.) *Computing Science and Statistics: Proceedings of the Twenty-First Symposium on the Interface*, 543-548 1989.

Campbell, G.; Douglas, M.A.; Bailey, J.J.: Nonparametric comparison of two tests of cardiac function on the same patients using the entire ROC curve. In: *Computers in Cardiology*, 267-270 1988.

Campbell, G.; Levy, D.; Bailey, J.J.: Bootstrap comparison of fuzzy ROC curves for ECG-LVH algorithms using data from the Framingham Heart Study. In: Kligfield, P.; Bailey, J.J.,(eds.), Computerized Interpretation of the Electrocardiogram XIV: *J. of Electrocardiology* 23, 1990 (in press).

Campbell, G.; Levy, D.; Lausier, A.; Horton, M.R.; Bailey, J.J.: Nonparametric comparison of entire ROC curves for computerized ECG left ventricular hypertrophy algorithms using data from the Framingham Heart Study improved method. In: Bailey, J.J.; Kligfield, P.(eds.) Computerized Interpretation of the Electrocardiogram XIII, *J. of Electrocardiology* 22, S152-157 1990.

Cropp, C.S.; Lidereau, R.; Campbell, G.; Champene, M.H.; Callahan, R.: Loss of heterozygosity on chromosome 17 and 18 in breast cancer carcinoma. In: *Proceedings of the National Academy of Sciences*, 1990 (accepted).

Hutchinson, G.: Admissible relations and coproduct congruences in universal algebra. At Jonsson Symposium on Algebra, Lattices and Logic, Laugervatn, Iceland, July, 1990.

Malley, J.: New methods for the statistical analysis of unbalanced or missing data problems. NICHD, NIH, Bethesda, Maryland, February, 1990.

Malley, J.: Statistical applications of Jordan algebras. In: "Lecture Notes in Statistics" series, Springer-Verlag, New York, 1990 (in press).

Malley, J.: The EM algorithm, Jordan algebras and Boltzmann learning machines; quantum statistical inference with applications to biomedical research. Department of Statistics, Harvard University, February, 1990.

Mosimann, J.: The statistical dependence of shape and size: size and shape variables with applications. Princeton University, Princeton, New Jersey, April, 1990.

Ratnaparkhi, M.V.; Mosimann, J.E.: On the normality of transformed Beta and Unit-Gamma random variables. Communications in Statistics—Theory and Methods (in press)

Senapathy, P.; Shapiro, M.; Harris, N.: Splice junctions, branch point sites and exons: sequence statistics, identification, and applications to the Genome Project. In: Doolittle, R. (ed.), *Methods in Enzymology* 183, 252-278 1990.

Willems, J.L.; Campbell, G.; Bailey, J.J.: Progress of the CSE diagnostic study—application of McNemar's test revisited. In: Computerized Interpretation of the Electrocardiogram XIV, *J. of Electrocardiology* 22, S135-S140 1989.

Wiseman, C; Gray, J; Mosimann, J.; Ahrens, A.: Cultural expectations of thinness in women: an update. *The International Journal of Eating Disorders* (in press).

CCB

Computer Center Branch

Joseph D. Naughton, Chief

The Computer Center Branch designs, implements, operates and maintains the central NIH Computer Utility and its associated telecommunications facilities. Two interconnected multicomputer facilities designed around large-scale IBM and Convex mainframe computers make up the NIH Computer Utility, which has evolved over two decades to provide state-of-the-art computational services in support of the dynamic and diverse requirements of NIH research investigators and administrators in the conduct and management of modern biomedical research.

The Utility provides interactive timesharing, data-base management, graphics, batch, high-performance scientific computation (e.g., vector), and word processing services to approximately 20,000 authorized users at NIH and in 29 agencies throughout the federal government as a Federal Data Processing Center (FDPC). All services are processed on a fee-for-service, full cost-recovery basis.

The IBM facility is made up of six IBM 3090 processors, each with one or more vector facilities, and an array of supporting peripheral equipment such as disks, printers, magnetic tapes and communications controllers. The six processors contain a total of over 768 million bytes of directly addressable memory. The peripheral complex includes 728 online disk drives with a total data storage capacity of over 1.2 trillion bytes; 108 cartridge tape drives with a transfer rate of 3 MB per second; 20 reel tape drives with a transfer rate of 780 KB per second, and four drives with a transfer rate of 200 KB per second; seven 18,000 lines-per-minute page printers, six 45 page-per-minute duplex cut sheet printers, two card reader/punches; and 13 telecommunications controllers supporting a teleprocessing network of over 1,500 lines of various bandwidths communicating with 330 remote job entry (RJE) computers and 10,000 interactive CRT and/or hardcopy interactive terminals located throughout the world.

The Convex system consists of a three processor Convex C230 minisupercomputer with 256 megabytes of memory, 15 gigabytes of disk storage, 4 magnetic tape drives, and 4 I/O processors dedicated to handling input and output from disk drives, tape drives, Ethernet connections, and terminals.

The two facilities are linked together by a high-speed network connection to facilitate the exchange of data, electronic mail, programs, and printing services. Ancillary equipment supporting both facilities includes two computer output microfiche units and film processors, two four-color high resolution plotters, and other miscellaneous devices.

Services of the Utility are available worldwide through the TYMNET, BITNET, and Internet international data communications networks as well as through the Federal Telecommunications Service (FTS), and commercial switched telephone services. Local area networks (LANs) are connected to the Utility through the NIH Computer Utility Integrated LAN and Workstation Support Network (NUnet).

The NIH Computer Utility operates 24 hours a day, seven days a week. It processes an average of 11,438 interactive sessions, 87,421 data-base transactions, and 18,000 batch jobs daily. Over 90 percent of all interactive commands are executed with subsecond response time, and six service classes for batch jobs guarantee maximum turnaround times of less than 30 minutes, one hour, two hours, and overnight.

Applications are programmed in FORTRAN,

COBOL, PL/1, WYLBUR Command Procedures Language, BASIC, Pascal, C, and Assembly Language, using three interactive user interfaces (TSO, WYLBUR and Convex Timesharing). Data-base management services are provided through DB2, a general-purpose, relational data-base management system, and IMS, a hierarchical, interactive transaction processing system for complex administrative applications. A variety of statistical analysis, modeling, data management, and utility programs are available, as well as the TELL-A-GRAF, POSTER and OMNIGRAPH interactive graphics packages. Molecular biologists and biological chemists have access to a number of relevant data-bases, such as GenBank (nucleic acid sequences), PIR (protein sequences), and PDB (protein structure).

In collaboration with several other DCRT laboratories and branches, the Computer Center conducts an in-house training program which offers 62 specialized classroom lecture courses in two semesters and a summer term. An interactive Assisted-By-Computer (ABC) training system permits users to develop expertise in the use of the Utility in their own work environment. In addition, 25 textbook courses are available for loan to students in a variety of topics. Technical documentation ranges from a comprehensive Computer Center Users Guide, which defines all available services, standards and procedures, to a technical newsletter, INTERFACE, which keeps users up-to-date on the current status of the Utility and announces changes and upgrades as they occur.

A highly experienced staff of professional, technical and management personnel provides assistance to users and ensures smooth functioning of the NIH Computer Utility. Computer specialists, systems programmers, and analysts develop and maintain operating systems software, provide technical consultation on program design and problem resolution, teach training courses, write technical documentation and design and implement new services and features in response to user needs. Computer systems technicians and operations personnel operate and maintain the Computer Utility hardware and telecommunications networks and provide user support services. The Computer Center Branch management team develops long-term program goals, conducts acquisitions, assists users with project planning, maintains the design integrity of the Utility, and responds to changing user demands.

The effectiveness of computers in the support of modern biomedical research is furthered by research and development projects conducted by the Computer Center systems staff.

NUnet—New Dimension in Connectivity

Communications and connectivity moved forward during the past year through DCRT's ongoing efforts to support electronic communications "across the hall or around the world." The scope of NUnet, the NIH Computer Utility Integrated LAN and Workstation Support Network, was expanded to provide interconnectivity among local area networks (LANs) in all NIH buildings, both on and off-campus, the Computer Center mainframe computers, and international data networks.

NUnet integrates the existing Computer Utility teleprocessing network supporting full screen and line-by-line interactive with the

rapidly developing LAN connections via a network of fiber optic and copper conductors. Some fifty NIH buildings in the Washington metropolitan area are being linked by NUnet, providing high-speed electronic communications among hundreds of LANs throughout the NIH. Outside users of the NIH Computer Utility who request connection to the network are served as well.

Exchanging reports, research documents, and other material electronically was facilitated through the NIH Computer Center's new mainframe mail gateway. This gateway can interface 3Com 3Plus and TCP/IP mail systems to WYLBUR mail on the IBM System 3090 mainframe, Convex mail, and the BITNET and Internet international network. The gateway enabled users on a 3Com 3Plus network to exchange mail with users on TCP/IP-based networks using the mainframe as a gateway. The software provides the necessary format and control conversion and routes mail.

The NIH Computer Center's mail gateway provides direct electronic mail connectivity between the IBM 3090 mainframe computers and the worldwide Internet network, and high-speed mail transmission between the IBM 3090 mainframes and TCP/IP and DECnet-based LANs connected to LANs on the Computer Center network. The most extensive high-speed network in the world, Internet encompasses millions of users at more than 120,000 hosts worldwide through three basic components—ARPANET, MILNET, and NSFNET—all using TCP/IP protocols. The Internet is used heavily by research scientists to exchange papers, data, and information relating to ongoing research and other topics of interest. This link to TCP/IP-based systems gives users of attached mainframe computers and LANs high-speed mail connectivity between themselves and to the international networks through one of the most comprehensive electronic mail transmission facilities anywhere.

The link to the Internet through TCP/IP-based systems gives the NIH research community easier access to the NCI Cray supercomputer located at the Frederick Cancer Research Facility. Researchers needing to access the Cray no longer need to travel to Frederick; users with access to a TCP/IP-based workstation can access the FCRF Cray through the Internet or via a direct high-speed communications line that has been installed between NIH and the FCRF Cray.

Improved connectivity led to the increased popularity of international discussion groups called LISTS which have proven enormously useful to BITNET users at the NIH Computer Utility. Over 200 LISTS have been established at computer centers on the BITNET and Internet networks to encompass such interests as computer science, technology, and ethics; the physical sciences; health and nutrition; pharmacology; various aspects of medicine, biology, and technology; and management of the BITNET network itself. LIST subscribers send and receive comments and discussions to the LIST via electronic mail. NUnet has expanded access to LISTS to NIH users in attached 3Com 3Plus networks.

An evaluation of Macintosh-to-mainframe connectivity began with the Computer Center offering of Level 3 support to VersaTerm, a dial-up terminal emulation product marketed by Synergy Software of Reading, PA. VersaTerm allows Macintosh users access to WYLBUR and TSO, DB2 and IMS via the Protocol Converter, and Convex; it is also Kermit compatible and

can be used to transfer files between the Macintosh and the mainframes.

The NIH Centralized Bulletin Board System (ENTER BBS) was made available to NIH Computer Center users on the 4th of July, 1990. This system was designed by Computer Center staff collaborating directly with a group of eight or nine technical leaders from the NIH Computer Utility user community. ENTER BBS was then developed, supported and maintained by the Computer Center.

ENTER BBS is unique in that it is a system that permits NIH users to create and moderate their own bulletin boards, without being computer professionals. It provides an easy-to-use umbrella system for numerous bulletin boards, each with all of the standard facilities—bulletins, files (for downloading), and electronic conferences.

ENTER BBS was designed and implemented to directly complement the Computer Center's critical communications role. Because of this, it represents a major advance in the communications and collaborating facilities available to the entire NIH. ENTER BBS is a state-of-the-art method for communications among committee members, between Research Grant administrators and their grants recipients, personnel offices and government employees, and technically oriented individuals interested in similar topics.

Convex Supercomputer Up and Running

Within 9 months of installation, the new Convex system had over 600 registered users, and the system was frequently supporting more than 30 active concurrent users. System utilization often reached 95% to 100% because of the large number of CPU-intensive jobs; and in mid-summer a third processor, additional memory, and more disk storage were added to the system to accommodate the growing demand.

The new processor, a Convex C230, has three tightly coupled central processors with an aggregate capacity of 75 MIPS (million instructions per second) and 150 MFLOPS (million floating point operations per second). The three processors provide parallel computing via Convex's hardware-based Automatic Self-Allocating Processor (ASAP) technology, which allows a single processor to request additional processors to execute portions of code that can run in parallel. The three processors share 256 megabytes of memory, a twofold increase in the original amount of memory on the system. Eight new disk drives, each with nearly 1 gigabyte (Gb) of storage, increased the amount of online data storage space from 8.5 Gb to 16.1 Gb; and an additional input/output processor (IOP) was added, bringing the total to four.

A month prior to installation of the new hardware, the system was configured with ConvexOS version 8.0, a major release of the Convex UNIX operating system and associated utilities. The new operating system offered major extensions in batch and tape queuing systems, added security features, accounting tools, and priority scheduling to manage parallel programs, improved throughput, and compatibility.

A new release of Convex C, the programming language that is an integral part of the UNIX environment, was installed this summer. This step made the Convex operating system fully compliant with POSIX, the Portable Operating System Interface for Computing Environ-

ments standard; conformed to the ANSI standard; added important extensions; and provided easy compatibility for users who are converting their programs.

UNIX System Evaluated For IBM Mainframe

The popularity of the UNIX operating system and its widespread acceptance in the scientific and engineering communities led the NIH Computer Center to install and evaluate a UNIX-type operating system, called AIX/370, on one of the NIH Computer Utility IBM 3090 processors, making it available to NIH researchers for an evaluation period at no charge. Compatible with both the System V version of UNIX from AT&T and the 4.3 version of Berkeley UNIX, AIX supports widely used communications protocols such as TCP/IP (Transmission Control Protocol/Internet Protocol); thus it can be linked with a variety of systems. As TCP/IP is commonly available on non-UNIX systems, connectivity is ensured outside the UNIX environment.

AIX/370 offers the Transparent Computing Facility (TCF), which allows workstations, micros, and mainframe computers to be clustered in a way that allows each user to access the best computer for performing specific tasks. TCF is completely transparent to the user, who may access the computing power of any computer in the TCF cluster without regard for which computer is actually executing the task.

In addition to the evaluation of AIX/370, the NIH Computer Center has also begun to examine Version 3 of AIX on IBM's RISC System/6000 platform. This evaluation will determine how the unique capabilities of the two platforms (i.e., 3090 mainframe and RISC 6000 workstation) can complement each other in a collaborative processing environment. Version 3 of AIX incorporates the features of both the System V version of UNIX from AT&T and the Berkeley version of UNIX. It includes the TCP/IP communication protocols and Sun Microsystems Network File System (NFS). The RISC 6000 comes configured for both Ethernet and Token Ring networks, and includes language support for C, FORTRAN, PASCAL and Cobol. Because of the general acceptance of the RISC 6000 in the marketplace it is expected that a wide variety of scientific applications will become available for the RISC 6000 during the next several years.

Equipment Evolution

Sweeping changes in the makeup of the NIH Computer Utility took place as the transition to new equipment was completed. Most IBM hardware components have been replaced or upgraded, the MVS/ESA operating system was installed and implemented, and numerous other software products have been made available to users. The result is a system that offers extremely economical computing. The evaluation has demonstrated the viability of the 3090 as a compute server and has led to the development of plans to investigate a collaborative computing network. Such a facility would include high-performance computer servers, file servers and visualization workstations working together to attack individual problems. This project is in the early planning stages but shows promise as a useful tool for scientific computing.

October 1989 saw the culmination of a 6-month effort to convert all the disk drives in the NIH Computer Utility to larger, faster, more

cost-effective model 3380K disk units. Nearly 500 volumes of double-density 3380E drives, containing over 350,000 data sets, were converted to triple-density 3380K drives. Also installed were 22 new IBM 3990-3 Storage Controllers. These replaced 32 IBM 3880-23 control units, providing additional functions and enhanced performance. These changes enable the NIH Computer Utility to provide over 1 trillion bytes of online disk data space to all six 3090-300E computational kernels.

Between January and March, the NIH Computer Utility's six IBM 3090-300E processors were upgraded to 3090-300J processors, which contain internal improvements that upgrade performance and provide increased cost-effectiveness. The size of the high-speed cache buffer was quadrupled from 64 KB to 256KB, and the execution time of the scalar multiply instruction was reduced 20%. Improvements to the vector facility increased its performance significantly, and the installation of faster chips reduced cycle time from 17.2 nanoseconds to 14.5 nanoseconds. The upgrades increased throughput by 30 percent and improved turnaround and system response time.

New standards of print quality became available to users of the NIH Computer Utility with the installation early this year of cut-sheet duplex laser printers. Providing single- and double-sided printing on 8-1/2" x 11" paper, the 3827 was greeted by a receptive user community. Use of the system grew so rapidly that additional printers were installed to handle over 3.5 million lines a day produced by 300-plus jobs. Six printers are currently handling the function with more planned, as additional capabilities, such as labels, will be offered on these printers.

New Software Products

The Genetics Computer Group (GCG) Sequence Analysis Software Package, a comprehensive set of programs for DNA and protein sequence analysis, was installed this year on the NIH Convex. GCG version 6.1, the most up-to-date version available, contains more than 125 programs and software tools for sequence editing, fragment assembly, mapping, comparison, data-base searching, multiple sequence analysis, pattern recognition, RNA secondary structure, protein analysis, translation, manipulation, display, and sequence exchange. Not only is the NIH Convex the first Convex system to have the GCG package, but users of the NIH Convex system are the first to utilize a supported, up-to-date version of the GCG package on a UNIX system. The GCG package runs in Convex's COVUE VMS emulation environment, which was significantly enhanced by Convex to support the package. Since the initial installation of the GCG package, graphic output, formerly limited to Tektronix graphics devices, has been extended to include all devices supported by the GCG Package, and an extensive set of printed documentation for the GCG Package has been made available.

Release 3.0 of SPSS-X became the production version of SPSS-X in February. SPSS-X release 3.0 offers a number of new and modified facilities and procedures, including five new subcommands for customizing currency formats and several other data formats, commands, and procedures.

Also in February, BMDP88 became the production version of BMDP. Two new programs are included in the new release along with a number of general enhancements.

Improvements and changes were made to both the DISSPLA and TELL-A-GRAF interface software to allow graphics users to take advantage of the IBM 3800 continuous-form and 3827 cut-sheet laser printers. The enhancements include a larger graphics area, sharper graphics picture, and three modes—draft, intermediate, and final—for output.

SAS/DB2 was made available under SAS Version 5.18 to provide interface procedures for transferring data between SAS data sets and DB2 tables. SAS/DB2 procedures allow users to extract data from DB2 tables and place them in SAS data sets; create and modify DB2 tables with data from SAS data sets; and enter, edit, or execute SQL statements to perform operations on a DB2 table.

Rates: Still Falling!

Economies of scale, efficient management, state-of-the-art hardware and software, and innovative users have made this the 22nd year of steadily falling rates at the Computer Center. In October 1989, the charge for all classes of batch processing services on the IBM System 3090 were reduced 7.4 percent, from $0.27 per machine unit to $0.25. The lower rate, which applied to batch services involving WYLBUR, DB2, the vector facility, managed storage, and the cartridge tape system, was expected to save users of batch processing on the IBM System 3090 about 1.5 million dollars per year.

The cost-effectiveness and reliability of the 3380K direct-access storage devices and 3990 control units obtained made an additional reduction possible in May 1990, when the cost for public online data storage was reduced 23%, from 1.3 cents to 1.0 cent per track-day. At the same time, a reduction was applied to the charge for IMS services which saved users of IMS an average of 28% per IMS transaction.

A new facility gave Computer Center users interactive access to billing data. Developed in response to numerous requests, PAS ONLINE provided a simple and efficient way for users to maximize computing efficiency and minimize cost by monitoring exactly where and how computing dollars are spent. With PAS ONLINE, users and account sponsors can display or print billing information for the previous full month, the current month-to-date, or a single day by using a series of simple-to-use DB2 Query Management Facility (QMF) procedures. Response to PAS ONLINE was enthusiastic, with users producing over 400 individual reports in the first two weeks the facility was available.

New Methods Revolutionize Training

Computer training opportunities increased explosively over the past year in order to keep up with the advances and changes experienced by the NIH Computer Utility. Forty-nine new courses and seminars were offered, and entirely new training methods were introduced in order to provide adequate training opportunities to all users.

A completely new computer-based training system, called Crwth (rhymes with truth), was made available this year. Crwth is a full-screen self-study system from Crwth Computer Courseware which enables users to take courses directly from a terminal or personal computer under TSO. Short, concise, and informative, Crwth courses are divided into sessions to give students the option of taking an entire course or

skipping to areas of specific interest. There are no online tests or quizzes.

Crwth courses supplement the Computer Center's first online training system, Assisted By Computer (ABC), which now has eight different courses. Because they are written by in-house staff, ABC courses take more time to write, test, and update. Crwth courses, written in general terms and kept current by the vendor, need only be modified to reflect the specific NIH operating environment and standards. Seven Crwth courses are available. They are:

- Using TSO at NIH
- Using ISPF/PDF: Fundamentals of Data Management
- Data Communications, Connectivity, and LANs: An Introduction
- Relational Data Base Overview: DB2
- SQL for Programmers: Coding DB2 Qeries
- Using ISPF: Advanced Features
- Using SAS/GRAPH.

ABC courses remained popular and filled an important need. More than 9,300 interactive ABC sessions were held during the year. All eight ABC courses were updated and new versions were made available.

During the spring semester, the DCRT Computer Training Program with input from other Division Branches and Laboratories, offered an expanded menu of formal classes and seminars in a wide variety of subject areas. Several new classes were added, including classes on UNIX, COMPRO utilities, and WYLBUR command procedures. Traditional mainframe and personal computer courses included courses in DB2, FORTRAN, COBOL, WYLBUR, SAS, Personal Computing, PC-DOS, and Batch Files with PC-DOS. A total of 210 sessions of 62 different courses was offered in the formal classroom training program. Enrollment in the formal training program was requested by 1,962 individuals, of which 1,868 (95%) were accepted.

Seminars provided substantial educational opportunities this year. Seminars included discussion groups on UNIX applications, an introduction to CONVEX, and workshops on mainframe connectivity that ranged from the basics of electronic mail to advanced topics for LAN administrators and technical personnel. Other seminars emphasized applying computing to research problems with such topics as "Remote Optical Sensing in Biological Tissue," "Image Processing on the PC," and "Recurrent Problems in Data Analysis." Experienced FORTRAN and C programmers learned how to use Convex's compiler technology more effectively at two seminars taught by Convex software support personnel. A total of 88 seminars were sponsored by the Computer Center last year, including 53 which were in the DCRT Computer Training Program or were presented upon request for interested groups at NIH, and 35 which were sponsored by the USAS for internal training.

Self-study training opportunities were increasingly important as users' needs grew. Several new self-study courses, which typically consist of a textbook, a workbook, and diskettes containing interactive exercises, were added this year to meet the burgeoning demand. New self-study courses dealt with DB2 Utilities, UNIX, and PL/1 programming. Two additional self-study courses featured topics in telecommunications and computer connectivity. "Individual Training for Local Area Networks," a tutorial, offered a brief introduction to the basic concepts and terms used in local area networks and PC-to-mainframe connectivity. PROTOCOL 90 pro-

Selected Courses and Seminars in the DCRT Computer Training Program Coordinated by CCB

Seminars
ENTERMAIL
Technology for Connecting Networks to Mainframes at NIH
Object-Oriented Programming with C++
Drawing DNA Sequences With Computers
Usage and Applications of Molecular Quantum Mechanical Programs
Using VAX VMS Resources at NIH

PC
Intermediate PC-DOS
PC-DOS Advanced Topics
Memory Management on the PC

Macintosh
Recovering Macintosh Disks and Data
Mac to DOS and Back
Macintosh Shareware

IBM 370
Advanced JCL at NIH
Introduction to ISPF/PDF
Designing Tables and Managing a DB2 Database
Use of Statistical Analysis Systems (SAS) at NIH

UNIX
Fundamentals of UNIX
Introduction to the X Window System
GCG Sequence Analysis
Convex Vectorization

vided a comprehensive study of telecommunications technology through 45 hours of interactive training on topics that included terminology, concepts, management, and practical and technical issues.

Another computer-based training course, "Fundamentals of the SAS Software System," was reviewed by the Statistical Software Section of the DCRT Laboratory of Statistical and Mathematical Methodology and made available online as part of the SAS software system. The course contained three instructional modules of two to six lessons each. Designed to provide an introduction to SAS, the course introduced only a fraction of the information presented in the formal training course, "Use of SAS at NIH." Up-to-the-minute information on the classroom courses, seminars, and self-study programs offered through both the DCRT Computer Training program and the NIH Training Center were made available online in October 1989. By using the ENTER TRAINING command in WYLBUR, users could access the latest news about training opportunities, including courses filled and new sessions scheduled, without waiting for office hours.

Documentation and Computer Assistance

Extensive new documentation was produced this year to describe the function and operation of the new Computer Utility hardware and software. Numerous new publications dealt with the MVS/ESA operating system and the Convex system.

There were 5 new technical documents published and 20 revisions or updates. The Automated Documentation Service was used by 4,800 users and 101,923 copies of technical publications were sent to those users and to individual requesters. INTERFACE, the Computer Center's technical newsletter, published six editions plus the annual index.

User assistance has become more important than ever in the dynamic environment of the NIH Computer Utility. To meet growing needs for consultation, the Computer Center increased the operating hours of the Consulting Desk in April by 66%. A popular service to thousands of computer users for nearly two decades, the senior computer specialists of the PAL Unit Consulting Desk now provide expert help throughout the business day, from 8:30 a.m. until 4 p.m. An integral part of the Computer Center's user assistance effort, the Consulting Desk is a vital part of the Computer Center's ongoing efforts to provide a wide range of assistance resources to users of the NIH Computer Utility.

Computer Center consultants handled 182 Consulting Appointments and 20,325 regular requests for customer assistance during the past year. In addition, 2,824 Programmer Trouble Reports were researched and answered. Changes to improve the performance and reliability of the operating system required the implementation of 24 SYSGENS (software reconfigurations). Approximately 23,200 fixes, both preventive and corrective, were tested and applied to the system and 31 new releases of current software packages were installed.

PCB

Personal Computing Branch

David C. Songco, Chief

The DCRT Personal Computing Branch (PCB) provides guidance and support to scientists and administrators throughout NIH in the development and effective use of personal computing. This support encompasses workstations, local area networks, and associated automation technology. The PCB assesses the technical requirements of the user community in the area of personal computing technology and collaborates with other DCRT groups to ensure that those requirements are reflected in the development and support provided by DCRT. Services provided by the PCB are based on the concept of establishing a partnership with the user community. This is accomplished by working closely with scientists and administrators and through a variety of user groups that focus on particular application areas and aspects of personal computing.

A staff of 22 computer specialists, engineers, and instructors provide consultation and problem resolution for a wide range of personal computing hardware and software products. PCB specialists develop and deliver training to enhance users' skills in personal computing and are constantly tracking the rapidly evolving distributed computing environment. PCB staff also create and distribute technical publications on personal computing to over 6,000 users at NIH.

The PCB is currently organized into four sections, each focusing on a specific area of technology. The *Administrative Technology Section* specializes in the technology of the PC family of computers. This section is also responsible for coordinating PCB publications and provides leadership in the area of computer security. The *Scientific Technology Section* currently tracks the Macintosh technology and leads the PCB support for biomedical research programs at NIH. The *Communications Technology Section* tracks the complex area of computer communications and focuses on the rapidly evolving area of local area networks. A fourth section, the *Consulting Services Section*, operates the PCB consulting service, provides guidance and support for PC data base and word processing applications, and coordinates the PCB Lead User Program.

PCB Helps Users Help Each Other

The PCB provides support and guidance for users of computing facilities owned and operated by the users themselves. To accomplish this, the Branch must enter into a partnership with the user to effectively provide its services. In this way, the PCB helps users help themselves.

This partnership is the basis for the highly successful PC Lead User Program established in FY84. More than 190 ICD-nominated NIH employees serve as PC lead users. These specially trained users collaborate with the PCB and serve as the first line of support within the NIH personal computing user community. Lead users answer PC-related questions, provide training, and disseminate PC-related information on behalf of the PCB.

A successful restructuring of the Lead User Program started with a "Lead User Kick-Off" in January 1990, which introduced lead users to the new format of the program and the certification process. Lead users actively participated in monthly Lead User Topic Sessions by demonstrating PC products, giving talks on PC-related issues, and sharing personal computing experiences.

Collaboration is also the basis for services provided to the rapidly growing Macintosh user

community. In FY90, PCB's Macintosh Support Coordinator Program was launched. Macintosh support coordinators are NIH employees who provide the first line of support to Macintosh users in their organization. Like their PC counterparts, Macintosh support coordinators share technical information and work closely with the PCB in NIH-wide Macintosh initiatives such as virus protection and program support. By the end of FY90, over 50 Macintosh support coordinators from various ICDs had enrolled in the program. PCB also continued to give full support to the Biomedical Research Macintosh Users' Group (BRMUG), a coalition of NIH Macintosh users. BRMUG provided a lively forum for the exchange of Macintosh information at NIH.

As yet another extension of the concept of helping users help themselves, this year the PCB also established the Campus Users Research Exchange (CURE). The CURE is a group of users representing local area networks across the NIH campus that meets monthly to exchange technical information. This year, the CURE focused on the PCB-supported 3Com network operating system and its role in the evolution of networking at NIH.

User Resources and Training

The NIH User Resource Center (URC), sponsored by PCB in collaboration with the Division of Personnel Management and the Division of Management Policy, is a multipurpose center equipped with both PC and Macintosh workstations and a variety of peripherals, including laser and matrix printers, plotters, and page scanners. The URC served as a repository and distribution point for PCB technical documents available for pickup by NIH employees, and offered self-study courses on personal computers and videotapes.

NIH employees made over 4,000 visits to the URC in FY90. These visits were used to take a tutorial, experiment with a particular type of equipment, pick up a document, or browse through personal computer magazines. The URC sponsored an effective Learning Assignment Program, where NIH employees worked in the URC for an agreed-upon number of hours each week in order to learn more about personal computer hardware and software.

The PCB-supported personal computer training program was expanded and improved in FY90. NIH scientists and staff enrolled in 245 sessions of 37 different courses. Vendors selected by PCB and the NIH Training Center presented 20 of the courses; the remaining 17 courses were taught by PCB, DCRT, or other NIH staff. Roughly a third of the PCB/DCRT courses were administered by the DCRT Training Unit and offered without fee. The cooperative arrangements between the NIH Training Center and DCRT functioned smoothly and productively throughout FY90.

Macintosh training expanded dramatically in FY90. Over 1,000 students enrolled in 100 sessions of 16 hands-on Macintosh courses available through the NIH Training Center, in addition to the lecture and/or demonstration Macintosh-oriented seminars given by the DCRT Training Unit. The number of these short DCRT Macintosh-oriented seminars will be substantially increased in FY91.

A vital factor in the high quality of PCB training is the PCB Associate Instructor Program, which illustrated the PCB philosophy of maintaining a close working relationship with

users by involving them in all aspects of support and guidance. About 80 NIH employees served as volunteer associate instructors for an average of five days each during FY90, making it possible for each hands-on class to have both a main instructor and two assistants.

Technology Tracking

The PCB constantly tracked and evaluated new products and technologies to ensure that the most appropriate and effective personal computing tools were presented to the user community. PCB staff found this the most challenging and enjoyable task they perform.

The PS/2, DOS, Windows, and OS/2 Environments

Advances in technology during FY90 meant increased performance for users with PS/2 systems based on the IBM MicroChannel Architecture (MCA). As in previous years, a variety of new PC applications and updated versions of currently supported software products for the DOS environment were tested and evaluated by PCB staff, lead users, and other specialists within DCRT. A DOS-to-OS/2 transition team was formed to test the new Operating System/2 (OS/2) extensively in order to determine its potential for use at NIH and relevant support issues. Later, another team was formed to evaluate Windows 3.0.

Versions of currently supported software designed to be run under OS/2 continued to emerge in FY90; however, a full complement of products will not be available until at least the second quarter of FY91. Of particular note was the delivery of Lotus 123/G, a powerful 3-dimensional spreadsheet program that exploits the graphical environment of OS/2. These application programs, as well as communication and data-base software designed for the OS/2 environment, promise increased capability and improved user interface, which should lead in turn to increased productivity at the workstation.

The release in May of Windows 3.0 captured the attention of the computer media and generated considerable interest among DOS users at NIH. In response, the PCB set about evaluating Windows and formulating recommendations.

Interest in presentation graphics and applications merging graphics with text grew throughout FY90. Peripheral devices, such as printers and image and text scanners, enjoyed a reduction in cost with increased speed and resolution, which resulted in PCB's evaluation and support for several new products.

PC-based products to aid in designing, generating, and filling out forms continued to advance in FY90. To respond to the widespread interest in this technology, the NIH Office Technology Coordinators initiated a subcommittee on the use of electronic forms at NIH. The DCRT Office Technology Coordinator, a member of PCB, served as technical advisor to this subcommittee. The group automated and distributed an electronic letterhead, which allows correspondence and letterhead to be printed simultaneously, to the NIH community with guidelines for its use. The committee focused on evaluating forms-generation software, with the intent of providing guidelines and recommending products to ensure consistency in NIH-wide electronic forms generation.

Advances in portable computer technology and increased numbers of requests for advice on product purchase in this category prompted

the PCB to announce support for portable and laptop computers from IBM and Compaq.

The emergence of many new "multimedia" products, which call upon textual, video, and audio technologies in presenting information, led PCB to begin experimenting with this technology. Support for several products in this area should be announced next year as demand increases and cost-effectiveness of the technology improves.

Macintosh
Support for three new Macintosh models—the Macintosh portable and the new IIci and IIfx models—was announced in FY90. The high-end IIfx added to the blurring of the traditional distinction between personal computers and scientific workstations. Other advances included the coming of age of 24-bit color and the announcement of PCB support for color scanning and printing. With the emergence of CD-ROMs oriented toward technical data in FY90, PCB began using this technology to reference Macintosh-related technical information. FY90 also saw the release of Apple's Unix product, AU/X Version 2, which PCB is now evaluating.

Communications
Currently the PCB supports the 3+Share network operating system from 3Com, a mature product that works well with both PC and Macintosh computers. In order to keep up with 3+Open, several 3+Open products were beta tested, including the current release, version 1.1, and the 3+Open Macintosh product. Both products will continue to be scrutinized, but neither is ready for use by the mainstream user community.

In addition, PCB has been evaluating a Public Network Utility, an isolated local area network used to test and evaluate the latest generation network hardware and software and distribute appropriate network tasks, utilities, and upgrades.

Computer Security
Responding to an increased need for educating the computer community about security issues, PCB appointed a Security Coordinator. The PCB Security Coordinator helped the DCRT ADP Policy Office plan and implement a Computer Security Awareness Day in March 1990. PCB staff also developed and distributed guidelines on how to prevent workstations from becoming infected by viruses, and arranged for an NIH site license for IBM's virus scanning program.

Dissemination of Information

Five issues of *PCBriefs,* the PCB's technical newsletter for NIH microcomputer users, were published and distributed to 6,500 personal computing users at NIH this year. PCBull, an electronic bulletin board service begun by the PCB in FY89, remained popular, averaging 1,000 calls per month. Accessible 24 hours a day from personal workstations via telephone communications, PCBull was an integral part of the PCB support program. It offered NIH users information, updates, and usage tips on PCB-supported products, as well as PCB utilities and publications and useful public-domain files for the PC and Macintosh. In addition, users could ask questions online as an adjunct to the PCB Help Line.

The Macintosh conference on PCBull was particularly active this year, with Macintosh users receiving access to timely Macintosh-related information, free software, and documents. For example, electronic memorandum

stationery was developed by PCB in multiple Macintosh word processing formats for NIH and individual ICDs and posted on PCBull.

FY90 saw the publication of the long-awaited *PCB Macintosh Product Information Guide,* with detailed information about PCB-supported Macintosh products and NIH resources. Macintosh computer support was also a subject of a DCRT/PCB poster session given at the annual NIH Research Day.

Consulting and Support Services

PCB provided a wide range of consulting assistance on a variety of personal computer-related issues to nearly 8,000 personal computer users in FY90. The most frequently asked questions were about networks, WordPerfect, Macintosh, and data-base management.

PCB Macintosh support expanded in FY90 in an effort to keep pace with the growing NIH Macintosh community. The staff averaged over 100 Macintosh consults a month.

The number of LANs at NIH supported by the PCB tripled this year to over 100. PCB worked closely with other labs and branches of DCRT and a number of other government agencies to provide wide area network connectivity nationally and internationally.

PCB also worked closely with the DCRT Network Task Group (NTG) and the DCRT Computer Center Branch (CCB) to keep the DCRT local area network running smoothly. This included active participation in the DCRT Connectivity Group and the DCRT Server Administrators Group, as well as monitoring the communications efforts of other labs and branches within DCRT and the ICDs.

PCB implemented and supported a DCRT-wide network fax gateway that allows text and graphic material to be faxed from the 3Com electronic mail system. This gateway supported local and international faxes and had Telex and MCI mail capability.

Future Plans

A major planning effort has been undertaken jointly by DCRT and the NIH Training Center to meet the challenge of providing for the ever-increasing amount and complexity of personal computer training needed at NIH. To this end, a number of changes in both computer courses and instructional staff are being planned for FY91. New courses will be added, existing courses will be modified or restructured, and there will be a shift in the balance of courses taught by in-house staff and those taught by outside vendors. This will allow the in-house staff to devote more time and energy to course development and instructional quality assurance; while more actual classroom instruction will be delivered by outside vendors using specific NIH course materials developed and maintained by the PCB.

Support for Windows and OS/2 will be phased in during FY91, while DOS support will continue. Support for the electronic forms-generation solutions recommended by the OTC subcommittee will also be announced. PCB will continue tracking the rapidly evolving personal computing technologies in order to provide guidance to the NIH community.

Apple is expected to introduce System 7 in FY91, which will require a major transition effort for NIH Macintosh users and the PCB. A new lineup of lower cost Macintoshes is also

expected, as well as Apple's Data Access Language, which gives flexible desktop access to host data bases.

The coming year will see an explosion of products available for network use in OS/2-based servers. A major challenge facing PCB in FY91 will be to ensure a smooth migration from 3+Share to 3+Open for users who need to upgrade. Support will be included for the TCP/IP protocol and XNS under 3+Open.

During 1990, 3Com focused more on a structured product line, addressing issues of multivendor, multiprotocol networking. Because 3Com seems to be moving toward network integration and away from development of network operating systems, PCB will most likely track the Microsoft LAN Manager as the strategic network operating system for many NIH users. In any event, PCB and other DCRT organizations and NIH network representatives will need to look ahead toward the next-generation network operating systems.

PCB is investigating the increasingly complex world of data-base management, in particular the client/server technology, which is one of the major computing issues of the '90s. We plan to work closely with the DCRT Data Management Branch and the DCRT Computer Center Branch to track and evaluate this technology from the personal computing viewpoint.

Support and guidance in the area of scientific computing will be increased again next year. This will include computing under the OS/2, Macintosh System 7, AU/X, and UNIX operating systems. Late this year a team was formed to evaluate the IBM RISC 6000 technology. In FY91 the PCB will continue this evaluation to determine if we should phase in some level of support for UNIX workstations.

Plans are being developed to restructure both the PCB organization and our consulting process to more effectively meet increased demands for support and guidance from the administrative and scientific community.

DMB

Data Management Branch

J. Emmett Ward, Chief

The Data Management Branch (DMB) provides advice and assistance to research investigators, program officials, and administrators throughout NIH in planning for and obtaining computer data processing services. A central NIH resource for systems analysis, design, and programming, the Branch also develops, maintains, and processes the NIH Administrative Data Base and the Clinical Center's Clinical Information Utility. On the staff are 44 permanent full-time employees whose disciplines include computer science, mathematics, and statistics.

DMB staff members design and create computer-based data management systems that provide practical solutions to the unique mix of administrative, scientific, and management data processing problems encountered at NIH. Comprehensive training in the facilities and functions of all systems developed by DMB is provided to each new computer system user. DMB staff members also teach courses about programming tools; provide advice on data management techniques to NIH programmers; serve as consultants to the Institutes/Centers/Divisions (ICDs) of the NIH for obtaining and monitoring contracting services for computer systems development; and create and maintain general-purpose, user-oriented programming tools for use in the development and operation of applications systems.

Three sections make up the DMB. The *Applied Systems Programming Section* (ASPS) and the *Scientific Applications Section* (SAS) provide general computer systems analysis and programming services to all of the ICDs, with ASPS supporting general data management and SAS handling projects that require scientific data analysis. The *Data Base Applications Section* (DBAS) develops and maintains the NIH Administrative Data Base for NIH material and financial management.

Broad Involvement in the NIH Mission

The NIH Administrative Data Base (ADB) represents a major effort by the NIH to integrate the administrative and financial data of its intramural program. A special approach to data base management, the ADB concentrates on the full sharing of data among all subsystems that support the NIH intramural program. It features on-line point-of-origin data entry, minimized data redundancy, background generation of all accounting transactions, and fully synchronized information processing (i.e., what you see is the latest state of any process, data or function).

The ADB is an ongoing developmental project. It currently supports these functions:

- Market Requisition
- Stock Requisition
- Central Procurement
- Delegated Procurement
- Central Stores Inventory
- Self Service Stores Inventory
- Biomedical Engineering and Instrumentation Program Inventory
- Division of Engineering Services Inventory
- Receiving
- Accounts Payable
- Source Contracts
- Vendor and Vendor Credit
- NIH Cashier Support
- Service and Supply Fund Activities (13)
- Travel Management.

DMB has continued the conversion of Service and Supply Fund activities to the ADB

during FY90. Software was completed to support the processing of all service requests to the Medical Arts and Photography Branch (MAPB). Software for special events and shipping/receiving/unloading was developed and made operational by the end of 1990. Design of software to support printing and reproduction is scheduled to begin shortly; and expansion of subsystems to support DES and transportation are planned for the coming fiscal year.

The first rendition of travel management, which included order entry, order change, and advance processing for domestic travel, was made an option last year in the ICDs. The processing and payment of domestic travel vouchers was tested during the last quarter of FY90, and this procedure will become generally available at NIH during the first or second quarter of FY91. Use of the travel ordering and vouchering system will become mandatory for domestic travel in FY92.

A mandate from the Office of the Secretary, DHHS, required NIH to implement fully automated, on-line fund control and fund certification procedures, along with a new government-wide standard general ledger which accommodates new accountability requirements. To meet this mandate, these processes are being added to the ADB. All of the relevant master control data files of the Central Accounting System (CAS) were converted for maintenance on-line in the ADB by the end of 1989. This year, DMB staff concentrated on developing the ADB facility that allows on-line distribution of funds from the appropriation level through the allowance level. The funds control software was tested during the last quarter of FY90. At the same time, procedures were developed for the fund certification process to occur at the allowance level with a warning message automatically queued to the Budget Officer if funds have been depleted. The procedures for fund certification and fund distribution should be implemented concurrently during FY91. Plans related to both fund control and fund certification slipped several months in 1989 when personnel were shifted to efforts related to the corrective action plan.

Processing for the government-wide Standard General Ledger (SGL) is in the definition stage. Plans are being made for all relevant NIH transactions to be directed to the SGL early in 1991. The the old and new ledgers may need to be maintained in parallel during the initial implementation of SGL.

The most significant ADB addition during the next fiscal year will be the implementation of the on-line Property Management System. Software has been completed to manage the day-to-day receiving, queuing and accounting for property items, and a major effort is underway to prepare for an NIH-wide property inventory and reconciliation. Inventories will be conducted by building; and as inventories are completed, procedures will be implemented to capture new property items based on relevant custodial codes for those buildings.

Early in FY90, the ADB maintenance contract was approved and a full complement of contractor staff was at work by the end of January 1990. By the end of the fiscal year, the contractor had responsibility for stock requisitioning, inventory management, delegated purchasing, source processing, receiving and accounts payable. Contract staff also worked with DMB to reorganize and update the ADB User's Guide; when completed, the guide will be maintained by the contractor.

The Cooperative Research and Development Agreement (CRADA) data base was fully implemented in early FY90, and CRADA information for NIH, CDC, and FDA was collected this year. As data were gathered, a data base was built which included information on each agreement, its principal and collaborating investigators, related inventions, patent applications, patents, and royalties. At the end of FY89, the Office of Technology Transfer (OTT) decided to process patents using a commercially available patent processing system. Patent royalties are currently processed by the National Technical Information Service (NTIS) in the Department of Commerce. Royalty data will also be maintained on the commercial system; and data related to ongoing CRADAs will be extracted for patents and royalties as soon as the commercial system is operational.

To support the Clinical Center Medical Records Department (MRD), a system was developed to control the retirement, location and reactivation of patient medical records. The system tracks the retirement of each medical record, monitors the microfilming process to ensure that administrators can control and evaluate contractor performance, establishes an easy locator method for retrieving archived records and assists in facilitating the reactivation process.

Developed during the 1970s as an historical archive of clinical information for research, the Clinical Information Utility (CIU) gathers data from the Medical Information System (MIS), the MRD and the various service organizations in the Clinical Center (CC). Over the years millions of records have been archived and made available for use in ongoing research protocols and for retrospective search and display. Over 95% of ad hoc requests are satisfied with overnight turnaround. The CC Information Systems Department monitors and authorizes all users of CIU data, and the CIU automatically tracks and reports each access of the data base. To satisfy clinical investigator needs, the CIU currently handles approximately 10 recurring and 20 ad hoc requests each week.

The CIU has now entered a maintenance phase. Efforts are centered on integrating CIU data classes and permitting overnight retrieval among them. Discharge diagnoses on the CIU for the years 1953 to 1986 were converted this year to the latest (ninth) revision of the International Classification of Diseases (ICD) codes. (Codes since 1987 have used the ninth revision.) DMB and Surgical Pathology staffs have also worked diligently this year to ensure that all surgical pathology data were kept current. Previously, Surgical Pathology data ran about one year in arrears.

DMB continued to support a limited service for analyzing nucleic acid and protein sequences using PC's; this included running data-base searches, advising investigators on different approaches for sequence analysis, and using a variety of commercial, public-domain, and tailor-made application programs to meet researchers' needs. Both the GenBank and the Protein Information Resource (PIR) were made available in the original format and DNASTAR format. Homology searches were supported via the Lipman/Pearson FASTA/TFASTA programs. Sequence-related work was supported by the MBUG programs and the DNASTAR sequence analysis package developed at University of Wisconsin. In addition to this ongoing effort, DMB also collaborated with the Laboratory of Mammalian Genes and Development, NICHD,

and the DCRT Physical Sciences Laboratory to create and maintain a data base of zinc finger proteins.

To attract highly qualified personnel to AIDS research, the NIH Office of AIDS Research is implementing a Loan Repayment program. This program offers participants who come to NIH the opportunity to have any loans which exceed 20% of salary paid by the NIH. DMB is developing a system to support the gathering of data on participants and their outstanding loans, the scheduling and review of loan payments and related taxes, the processing of each payment, the tracking of each payment to the appropriate treasury schedule, and the generating of IRS 1099 forms at the end of each tax year.

DMB also provided software support to report on AIDS research expenditures in the ICDs, using the Mason category codes and special interest codes. Quarterly reports of expenditures were generated, with details of this information stored in an on-line relational data base to support ad hoc queries. Input of expenditure data was received by way of different media (PCs and mainframes) and different modes (mainframe files, SYMPHONY data bases and LOTUS 1-2-3 data bases). Data were converted to a standard mainframe format, summary reports were generated, and a relational data base of the detail was created for on-line query using SQL. This software has been turned over to technical staff of the Office of AIDS Research, and DMB staff have agreed to remain available to the project for changes and enhancements through the end of the year.

DMB provided ongoing support, analysis, design and maintenance for an additional 37 projects during the past year. A brief overview follows:

- **For NICHD:** DMB staff conducted and presented an extensive requirements analysis to enhance CHIPS (Child Health Information Portfolio System); developed retrieval and report software for extramural grants and contracts, and trained NICHD personnel.
- **For the Office of Program Planning and Evaluation, OD:** DMB staff processed the Case reports to the National Science Foundation for FY89, which summarized DHHS-wide grant and contract expenditures by institutions of higher education, by research and development centers managed by institutions of higher education, and by non-profit institutions.
- **For NCI:** DMB continued supporting the Cancer Information Service, the Seroepidemiology Data Analysis, the Grants Literature System and the Grants Elemental Network Internal User System.
- **For DRG:** DMB supported the NIH Consultant File of scientific qualifications and the classification of NIH research and development awards.
- **For NHLBI:** DMB continued to support the cardiology/cardiac surgery data base.
- **For NIAID:** DMB continued the full time equivalency management system.

Future Plans

As the '90s begin, DMB will examine and evaluate its role in providing computational support to the NIH community. DMB will not only reevaluate its mission, but will modify the types of support that it offers to the NIH community. To accomplish these goals, the DMB will undergo a reorganization early in FY91. A New Technology Analysis Section will be established.

This section will be responsible for evaluating and recommending new technologies as they relate to data-base management, data retrieval, data analysis, and information processing; selecting new technologies for long-term support and use within the DMB; training DMB personnel in new technologies selected via internal evaluations or identified as essential to keep pace with accepted industry standards and strategies; and acting as an advisory group to others at the NIH concerning results of evaluations and proper use of selected facilities. This new section will produce the following results:

- new direction for the Administrative Data Base with either supplemental management information software or a totally new data-base management approach
- definition of and software support for data-base management across computing platforms
- better understanding of other types of support and their interrelationships with data-base and distributed support mechanisms, including graphical user interface, spreadsheets, graphics and statistics
- awareness of the marketplace and a good understanding of potentially successful trends in software and hardware support
- training of DMB and other NIH technical personnel in applications of highly useful software products.

Discrimination Prohibited: Under provisions of applicable public laws enacted by Congress since 1964, no person in the United States shall, on the grounds of race, color, national origin, handicap, or age, be excluded from participation in, be denied the benefits of, or be subjected to discrimination under any program or activity (or, on the basis of sex, with respect to any education program or activity) receiving Federal financial assistance. In addition, Executive Order 11141 prohibits discrimination on the basis of age by contractors and subcontractors in the performance of Federal contracts, and Executive Order 11246 states that no federally funded contractor may discriminate against any employee or applicant for employment because of race, color, religion, sex, or national origin. Therefore, the Division of Computer Research and Technology must be operated in compliance with these laws and Executive Orders.

Division of Computer Research and Technology
National Institutes of Health
Bethesda, Maryland 20892