Human Genome news

National Center for Human Genome Research
National Institutes of Health

A Forum For Human Genome Information Exchange

ISSN: 1050-6101 Vol. 2, No. 2, July 1990

DOE Human Genome Program Developing Sequencing Technologies

Evolutionary Improvements and Revolutionary Methods Development Under Way

Marvin Stodolsky, Human Genome Program, U.S. DOE

The primary goal of the Human Genome Project is to produce a reference DNA sequence. Although sequencing technologies have recently improved, sequencing the 3-billion-bp human genome will require faster, less expensive, and more accurate systems. The sequencing efforts being funded by DOE can be divided into two categories – evolutionary improvements and revolutionary methods. (See box on p. 2 for information on requesting copies of research abstracts of work funded by DOE.)

Those methods categorized as evolutionary improvements primarily support Sanger and Maxam-Gilbert sequencing strategies, or their extended versions, in which four families of DNA fragments are produced, each family ending in one of the four bases (A, T, C, or G). Size fractionation by gel electrophoresis (see photograph) produces a sequencing ladder with steps of increasingly longer fragments. The sequence is read as the end base on successively longer fragments within the four ladders.

In contrast, those methods categorized as revolutionary are innovative and high-risk. They rely on novel data-acquisition methods and promise to generate sequence data at higher rates, but they will require substantial technical development before sequence production begins. Both evolutionary and revolutionary systems are undergoing or will undergo validation trials to determine their accuracy, speed, and total-system economics.

Evolutionary Sequencing Improvements

Increasing Resolution and Detection Sensitivity

For current gel electrophoretic fragment separation systems, the capacity for spatially resolving successively longer fragments is diminished at lengths of several hundred base pairs. This decrease in resolution limits accurate reading of sequence from the sequencing ladder. Use of capillary or ultrathin slab

In This Issue...

Mapping and Sequencing

Abstracts of all projects mentioned are contained in the DOE *Human Genome 1989–90 Program Report*, except for newly funded sequencing projects on transposon-aided sequencing, an improved CCD optical system for flow cytometry, exciton microscopy, and coherent X-ray crystallography of DNA fibers. Copies of the program report and these abstracts may be obtained from HGMIS (address on p. 20).

gels offers promise that higher resolution will extend sequence "reads" above a thousand bases. Thin gels also rapidly dissipate ohmic heating and allow use of higher voltages to result in faster separation. A tenfold increase in fractionation speed is expected, with attendant increased DNA throughput.

To visualize the sequencing ladders, a variety of DNA labels have been used, including radioisotopes, stable isotopes, fluorescent compounds, and chemiluminescent materials. Replacement of radioisotopes with less hazardous labels is highly desirable. Labels may be attached to DNAs being fractionated, or they may be attached to oligomers (short, synthetic DNAs), which can then serve as probes that will selectively bind to their complementary sequences in a ladder and thus display their positions.

For sequences labeled by radioisotopes or chemiluminescence, the ladder images are first captured on film. Several versions of scanning densitometers and charge-coupled display (CCD) cameras are being evaluated for conversion of images into computer-manipulable data.

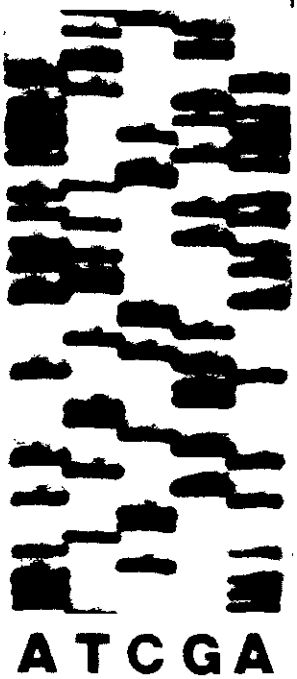
In support of systems using fluor labels for either mapping or sequencing, more sensitive instruments for detecting in-gel fluorescence are being developed. Decreasing the concentration of DNA in a gel provides some increase in resolution of successive fragment bands. Sensitivity gains are particularly important for analyses on capillary or ultrathin slab gels with their reduced DNA loading capacities.

Stable isotopes, four or more per element (e.g., tin), are particularly promising for multiplex sequencing or blotting procedures. When the isotope is incorporated into the primer used in Sanger or polymerase chain reaction (PCR) procedures, the electrophoretically produced DNA bands are located by scanning the gel with a resonance ionization spectrometer coupled to a time-of-flight mass spectrometer. The A, G, C, and T Sanger fragments, each labeled with an individual isotope, are run in a common gel lane; such fragments from several DNAs can also be combined into that same gel lane, as long as they carry distinguishable stable isotopes. Analysis rates of 3000 to 10,000 DNA bands per minute are suggested.

DNA fragments separated by gel electrophoresis. The bases are read according to their band positions in the ladders. The sequence of the top 15 nucleotides is:

TCGTGGTTAATG.

Photo provided by Beth Mullin and Paula Goetting-Minesky, The University of Tennessee, Knoxville.



Sequencing Strategies for Extending Reads

Today, physical maps of chromosomes are being generated as ordered libraries of cosmids or yeast artificial chromosomes (YACs). Several different strategies can be used to obtain the sequence of these large cloned DNAs when unit reads are only several hundred base pairs long.

In directed strategies, an initial mapping effort is necessary to identify suitable sites of sequencing initiation. They are chosen so that unit reads of DNA will overlap and permit extended sequence assembly by aligning the overlaps. In shotgun methods, initiation sites are randomly chosen. More total sequencing is necessary to guarantee overlaps of unit reads. There are also hybrid strategies that begin with shotgun sequencing and finish through some directed sequencing. PCR can serve to amplify and identify DNA regions needed for completion of an extended sequence.

Multiplex sequencing promises considerable increase in efficiency, because target DNAs are processed in pools of 20 or more sequences. Each individual member of a pool has a distinguishing "tag" sequence at the beginning of its sequencing ladder. Several processing steps culminate in the binding of fractionated fragment patterns to a membrane. Thereupon, each superimposed sequencing ladder yields its discrete sequence data as hybridizations are performed with oligomer probes complementary to each distinguishing tag.

Preparation of DNA for Sequencing

The cost of generating an extended sequence resides mainly in its preparation (e.g., reduction of cosmids or YACs into needed small recombinant DNAs, restriction fragments, or PCR fragments). Reduction of "front end" costs is thus important for total system economics. Automation in clone management, subcloning, and DNA preparation is progressing. Some newer sequencing strategies do not require the production of subclones from cosmids or perhaps even from YACs.

Transposable genetic elements with features that facilitate mapping and sequencing are being constructed. Within host bacteria, the transposon randomly integrates at many sites within the human DNA cosmid

inserts, and therefore many initiation sites are provided for sequencing. The DNA segments between integrated transposons must first be isolated so that each segment provides a unique sequencing template. Sequences generated from multiple insertion sites should overlap and provide sequence coverage for the entire human segment of the cosmid.

Similar cosmid coverage is achieved by using a family of oligomers to provide primers for Sanger sequencing. Each sequencing run identifies oligomer family members that can next be used as primers, so that spans of contiguous sequence are progressively extended. Eventually the spans overlap, and the sequence of the human segment within the cosmid is completed.

Data analysis is a major bottleneck in all contemporary sequencing projects. The read accuracy of sequence data varies; the shorter the fragment being read on the ladders, the greater the accuracy of sequence data obtained. This data must be computationally processed to recognize overlaps of sequence reads so that extended sequence can be assembled. This assembly task is currently being addressed by development of special algorithms. The output provided for refined displays of probable overlaps of reads also aids and speeds any necessary human decisions.

Revolutionary Sequencing Methods

Revolutionary sequencing methods are diverse in operational principle and instrumentation employed. The DOE-funded projects described below demonstrate this diversity.

Sequencing by Hybridization

This technology uses a unique strategy and various computer algorithms to assemble extended sequence from very short sequences. Some 100,000 oligomer probes are used to test for the presence of complementary sequences in recombinant DNAs of the single-stranded virus M13. The library of M13 recombinants has a manyfold representation of the region to be sequenced. Short sequence data obtained with probes first serve to order overlapping clones of the library; then extended sequence is assembled from the short sequence data. When false branchings in sequence within a single clone do arise, they are eliminated by comparing results from several clones that partially overlap the branch region. Proof-of-concept sequencing has been performed successfully, with a limited family of oligomers, on a 100-base-long M13-interferon DNA.

Sequencing with Flow Cytometry

Some revolutionary approaches seek to achieve sequence acquisition from a single strand of DNA. The successive-base-release approach utilizes automated flow cytometry technology. In a preparative step, one strand of a duplex is partially degraded and then replaced, through DNA polymerase action, with base subunits labeled with distinguishing fluors. The labeled DNA is mounted in a quartz capillary tube, together with a suitable processive exonuclease. Bases sequentially released into the flow stream would then be identified by their distinguishing fluor labels. A detection system capable of detecting and identifying single molecules is essential to this effort.

Sequencing with Scanning Tunneling Microscopy (STM)

Since the January 20, 1989, *Science* report on STM of DNA, this sequencing approach is being pursued in many laboratories. The report resulted from a collaboration between Lawrence Berkeley Laboratory (LBL) and Lawrence Livermore National Laboratory (LLNL).

In STM technologies a tip approaching atomic dimensions scans the specimen surface nondestructively. Atom-scale resolution of simple surfaces has been achieved with this approach. With DNA specimens, the objective is to distinguish the four bases on the sugar-phosphate backbone of DNA.

The newest member of the scanning microscopy family is molecular exciton microscopy. In this technology, the energy exchange and modulation between a near-field optical scanning tip and specimen is sensitive

DNA Sequencing Literature Collection Available from USB

The United States Biochemical Corporation (USB) has announced the availability of its DNA Sequencing Literature Collection, which consists of reprints of popular journal articles on the enzymology or theory of DNA sequencing and on DNA sequencing methodologies. The collection also includes issues of *Comments*, the USB quarterly newsletter, which contains technical articles about new products and programs. Updated and supplemented regularly, the collection is available at no charge. Make requests for these publications to United States Biochemical Corporation; Box 22400; Cleveland, OH 44122; (800) 321-9322, in Ohio: (216) 765-5000. ◇

Mapping and Sequencing

Microimaging Workshop Proceedings Available

X-Ray Microimaging for the Life Sciences (NTIS: DE 90002613), a 212-page book of proceedings consisting of extended abstracts from the workshop held May 24–26, 1989, in Berkeley, California, is available for $34.00 from the National Technical Information Service; U.S. Department of Commerce; 5285 Port Royal Road; Springfield, VA 22161; (800) 336-4700.

to the detailed nature of electronic orbitals in the specimen. For DNA analyses, differences between bases would be enhanced by attachment of distinguishing metal labels.

Sequencing with X-Ray Lasers

Recent development of intense, coherent (laser) X-ray sources and high-quality X-ray optics may result in X-ray microimaging capabilities with sufficient spatial resolution to define base sequence. In principle, a single strand could provide sufficient data to reconstruct a holographic image of the molecule. Very high performance and a means to enhance contrast between the bases would be essential for all subsystems. Radiation damage to the DNA will be a particular problem to address.

In a less demanding quasi-crystallographic approach, a chromosome segment is amplified by PCR, and one of the base types labeled with a heavy metal. Samples containing as few as 10 million labeled and fully extended DNA molecules might suffice as a target. Theoretical analysis shows that by using a coherent X-ray source, the positions of metal labels, and hence bases, on the DNA fragment can be determined from scattering data. For each chromosome segment, four fibers corresponding to individual labeling of A, T, C, or G would be necessary. Sequence would then be obtained by integrating the four sets of position data.

NIH-DOE Implementation of Strategies

With many evolutionary improvements in progress and with the revolutionary schemes now under development or not yet imagined, prediction of future genome sequencing technologies is difficult. To better assess progress and coordinate research in the sequencing portion of the nation's Human Genome Project, the DOE Human Genome Program and the NIH National Center for Human Genome Research have organized the Joint Working Group for DNA Sequencing (see related article below). ◇

Joint Sequencing Working Group To Study Research Priorities

The human genome programs of NIH and DOE have established the Joint Sequencing Working Group to study and make recommendations on research priorities needed to meet the goal of sequencing 3 billion nucleotides of human DNA within 15 years. Although recent improvements in sequencing technologies are reducing the time and cost required, continued improvement is needed before scientists will be able to sequence the entire genome.

Joint Sequencing Working Group

Ellson Chen	Genentech, Inc.
Ronald Davis	Stanford University
John Devereux	Genetics Computer Group, Inc.
Walter Gilbert	Harvard University
Leroy E. Hood	California Institute of Technology
Mark Pearson	E. I. du Pont de Nemours & Co.
Joseph Sambrook*	University of Texas Southwestern Medical Center
Phillip A. Sharp	Massachusetts Institute of Technology
William Studier	Brookhaven National Laboratory

*Chair

Within 5 years, the Human Genome Project seeks to improve existing methods or develop new ones to lower sequencing cost from the current $3 to $5 per base pair to well below $1. Another 5-year goal is to sequence an aggregate of 10 Mbp of human DNA (and 30 Mbp from model organisms) in large, continuous stretches.

To date, complete DNA sequences have been determined for some viruses, of which the largest sequenced so far is the cytomegalovirus containing 240 kbp. Investigators are attempting to sequence bacterial genomes that contain about 4.5 Mbp. Although many short stretches (an aggregate of nearly 7 Mbp) of human DNA have been sequenced, the complete sequence of the human genome will total some 500 times that amount. Projects will be initiated to develop and assess strategies and technologies for sequencing whole chromosomes and for reducing the cost.

To begin formulating strategies for achieving these goals for sequencing, the working group met on May 10 in Herndon, Virginia. ◇

Reported by Leslie Fink, Chief
Office of Human Genome Communication
NCHGR

Whose Genome Is It, Anyway?

Leslie Fink

NIH National Center for Human Genome Research

A quick glance around any public gathering will attest to the physical diversity of the human population. In most groups of people, some will be tall, others short; some will have brown eyes, and others blue. These physical attributes are largely determined by genes – packets of the chromosomal genetic material, DNA.

The complete human DNA, collectively called the human genome, is made up of about 3 billion chemical subunits called nucleotides. The task of the Human Genome Project is to map all the genes on human chromosomes and to determine the order, or sequence, of the 3 billion nucleotides. As scientists begin to map and analyze the molecular details of the human genome, the question is often asked, Whose genome is being used?

Are scientists using just one genome?

In many ways, describing the anatomy of the human genome will be similar to studying the human heart or the human brain. While there are small differences from person to person in the size and shape of these organs, most key characteristics are the same. "Although human beings are distinct from one another, they are really very similar in most biologically important respects," says Mark Guyer, Assistant Director for Program Coordination, National Center for Human Genome Research (NCHGR). "That's what makes us human. So the map of the human genome can really be based on information collected from many different people. And most of the information in that map will pertain to everyone."

Geneticists estimate, for example, that any two people are about 99% similar in their genetic makeup. The tiny differences between any two people rest in only 2 to 10 million (out of the 3 billion total) nucleotides, an amount that computes to less than 1% of their total DNA. " Because these small differences vary from person to person," says Guyer, " it doesn't matter whose genome it is." Furthermore, it doesn't matter how many different genomes are used.

How will various genomes be integrated into one reference map and sequence?

Eventually, scientists will "map," or establish distinctive genetic landmarks, from one end of a chromosome to the other and add that information to the genetic map of the entire human genome. This complete map will become the reference to which researchers will compare DNA taken from a variety of people as scientists look for disease genes and other important genetic regions located on chromosomes. A particular region on a chromosome, for example, may be found to contain information about height. Although the genetic content of that specific site may change slightly from person to person, the location of the site will be the same in each person's genome.

Because studying the entire 6-foot stretch of human DNA is a huge project, scientists are tackling the genome one chromosome at a time. Nevertheless, analyzing the genetic content of just one chromosome is an enormous task for a single research group, so many scientists are studying portions of chromosomes. The complete map for a single chromosome will, therefore, be derived from samples collected by researchers in many different laboratories from unrelated individuals.

David Ledbetter and his colleagues at the Baylor College of Medicine, for example, are mapping the sex chromosome X and chromosome 17. "We collect DNA from patients who come into our clinic for genetic testing," says Ledbetter, "so each sample is from a different and unrelated person. Our cell culture collection contains a number of different human genomes."

According to David Schlessinger, Director of the Center for Genetics in Medicine at Washington University in St. Louis, "The ultimate goal of the genome project is to have the technology and ability to check parts or all of many genomes. We will analyze one genome relative to another, and that's part of the interest in this project."

Referring to the donor of the X-chromosome region Schlessinger's group is studying, he states, "The identity of the donor is locked away . . . . that individual is the genetic equivalent of the unknown soldier." ◇

Human Genome news

NATIONAL INSTITUTES OF HEALTH

This newsletter is intended to facilitate communication among genome researchers and to inform persons interested in genome research. Suggestions and contributions are invited.

Managing Editor
Betty K. Mansfield

Editors
Anne E. Adamson
Judith A. Aebischer
Judy M. Wyrick

Production Manager
Judy M. Wyrick

Production Assistants
Larry W. Davis
Michael D. Mayes
Laura N. Yust

Advisors
Leslie Fink, NIH
Jeffrey E. Schmaltz, DOE

Special Thanks to:
Charles R. Cantor
Anthony V. Carrano
Leroy E. Hood
Kathleen H. Mavournin
John S. Wassom
Brad L. Whitfield

Correspondence Address:
Betty K. Mansfield
ORNL
P.O. Box 2008
Oak Ridge, TN 37831-6050
Phone: (615) 576-6669
FTS 626-6669
Fax: (615) 574-9888
FTS 624-9888
Bitnet: "bkq@ornlstc"
Internet: "bkq@ornl.gov"

DOE Program Office
Germantown, MD
Phone: (301) 353-5037
FTS 233-5037
Fax: (301) 353-5051
FTS 233-5051

National Center for Human Genome Research
Bethesda, MD
Phone: (301) 402-0911
Fax: (301) 480-2770

DEPARTMENT OF ENERGY · UNITED STATES OF AMERICA

Genome Project News

HUGO News

Grants Help Support New Regional Office

HHMI and Wellcome Trust Announce Awards to HUGO

The Human Genome Organisation (HUGO) has received funding grants from the Howard Hughes Medical Institute (HHMI) and The Wellcome Trust of the United Kingdom to cover operating costs of its new American and European offices.

In May, HHMI announced the award of a $1-million, 4-year grant that will partially fund operations in the HUGO Americas office in Bethesda, Maryland, and its related international activities. These include some meetings of the HUGO Council and its committees and contributions toward a scientist-exchange program.

Earlier this year, The Wellcome Trust in the United Kingdom announced a 3-year grant to HUGO to assist with operations of the European office in London and HUGO's activities in Europe. The Wellcome Trust is providing approximately $296,000 in the first year, with an additional $84,000 available for program activities such as single-chromosome workshops. Subsequent allocations are expected to be smaller as HUGO succeeds in obtaining longer-term support from governments. Probable location of the HUGO London office will be in the headquarters of The Wellcome Trust.

HUGO Objectives and Membership

Major objectives of HUGO are to foster international collaboration among scientists, to serve as a coordinating body in the international Human Genome Project, and to help coordinate physical mapping of individual chromosomes. To accomplish these goals, HUGO will establish international training programs on relevant methodologies and will facilitate the exchange of appropriate data, samples, and technology. HUGO also plans to foster parallel studies of model organisms, such as the mouse, and to coordinate research with the U.S. Human Genome Project. HUGO is committed to encouraging public discussion about the societal impact of data derived from the genome project.

Formed in 1988, HUGO has an elected membership of 239 prominent scientists from 23 nations. U.S. members of the HUGO Council serve to coordinate its activities with U.S. federal agencies, especially NIH and DOE, leaders of the U.S. effort. HUGO's first president was Victor McKusick (Johns Hopkins Medical School), and the second and current president is Sir Walter Bodmer (Imperial Cancer Research Fund, London). See box for list of HUGO officers.

HHMI Involved Directly in Mapping and Sequencing Research

Through its research program, HHMI has been extensively involved in mapping and sequencing human genes. Since 1985 HHMI has supported collection and dissemination of information on genome mapping through a network of databases, including the new genome data library and information center supported by HHMI at Johns Hopkins University (see related article, p. 16).

HHMI, established in 1953, employs scientists in the fields of cell biology, genetics, immunology, neuroscience, and structural biology. HHMI investigators conduct their research in collaboration with outstanding academic medical centers and universities throughout the United States. The Institute also supports science education through its grants program.

Wellcome Trust Supports Medical Research

The Wellcome Trust, Britain's largest medical research charity, was established in 1936 under the will of Sir Henry Wellcome. The Trust derives its income almost entirely from its investment portfolio, which includes 75% of the share equity of Wellcome PLC, the multinational pharmaceutical enterprise operating in North America as Burroughs Wellcome.

The Trust supports all aspects of medical research (except cancer), primarily through grants to investigators in British universities and medical schools. A total of

(see HUGO, p. 7)

HUGO Officers

President

- **Walter Bodmer [Imperial Cancer Research Fund (ICRF), London]**

Regional Vice Presidents

- **Charles R. Cantor, North America (Director, DOE Human Genome Center, Lawrence Berkeley Laboratory)**
- **Kenichi Matsubara, Asia (Director, Institute for Molecular and Cellular Biology, Osaka University, Japan)**
- **Andrei D. Mirzabekov, Eastern Europe (Director, Institute of Molecular Biology of the Academy of Sciences in the Soviet Union)**

Secretary

- **Bronwen Loder (ICRF)**

Treasurer

- **George Cahill (Howard Hughes Medical Institute)**

Galas Appointed New DOE Associate Director

David J. Galas assumed his duties in April as the DOE Associate Director of the Office of Health and Environmental Research (OHER). He is responsible for OHER's five divisions, with a budget of $300 million: Health Effects Research Division (including radiation biology, molecular biology, and general life science research); Human Health and Assessments Division (including nuclear medicine and epidemiology research); Physical and Technological Research Division (including scientific instrumentation and radiation detection and measurement); the Ecological Research Division; and the Atmospheric and Climate Research Division. Major interdivisional programs for which he is responsible include Human Genome, Structural Biology, Radon, and Subsurface Biology.

Former Director of Molecular Biology and Professor of Biological Sciences at the University of Southern California, Galas received his B.A. in physics at the University of California at Berkeley. He was awarded the M.S. and Ph.D. degrees in physics by the University of California at Davis in cooperation with Lawrence Livermore National Laboratory (LLNL).

Before joining the faculty at the University of Southern California in 1981, Galas spent 4 years in the Molecular Biology Department of the University of Geneva in Switzerland. From 1974 to 1977 he was a senior staff scientist in the Biomedical Division of LLNL, and from 1972 to 1974 he served as scientific advisor to the Defense Science Board Task Force on Strategic Vulnerability.

His research interests have included the study of the transposition of genetic elements, the biological significance of DNA transposition in living cells, and the study of DNA-protein interactions. He has developed several techniques used in molecular biology research, including the widely used DNA "footprinting" method. ◇

David J. Galas
DOE Associate Director
Office of Health and
Environmental Research

DHHS Secretary Signs NCHGR Advisory Council Charter

Secretary of the Department of Health and Human Services (DHHS) Louis W. Sullivan recently signed a charter giving the National Center for Human Genome Research (NCHGR) authority to establish a national advisory council. Members will be appointed in the next few months.

As announced in the April 24 *Federal Register*, the National Advisory Council for Human Genome Research will serve in an advisory capacity in the conduct and support of the NIH Human Genome Program and in the dissemination of information about human genome research, training, and other related programs. The Advisory Council will also evaluate and recommend approval of applications for research grants, training grants, and cooperative agreements.

The Advisory Council will consist of the NCHGR director, who will serve as chair, 12 expert active members, and 3 other nonvoting ex-officio members. Of the 12 experts, 8 will be selected from the fields of basic genetics, medical genetics, molecular biology, biochemistry, information science, mathematics, and engineering; 4 will have backgrounds in public policy, law, ethics, and economics. ◇

HUGO *(from p. 6)*

$100 million will probably be allocated during FY 1989–90.

Director of the Trust, P. O. Williams, views the establishment of HUGO and its international coordinating role as a positive step in ensuring that all nations share in efforts to map and sequence the human genome. According to Williams: "The potential benefits accruing from this project in relation to understanding human diseases cannot be overestimated." ◇

Reported by Anne Adamson and Judy Wyrick
HGMIS, ORNL

This newsletter is prepared at the request of the DOE Office of Health and Environmental Research and the NIH National Center for Human Genome Research by the Biomedical and Environmental Information Analysis Section of the Health and Safety Research Division at Oak Ridge National Laboratory, which is operated by Martin Marietta Energy Systems, Inc., for the U.S. Department of Energy, under Contract DE-AC05-84OR21400.

Genome Project News

DOE Human Genome Coordinating Commitee

The DOE Human Genome Coordinating Committee (HGCC), which was formerly called Human Genome Steering Committee, held its fifth meeting in Pasadena, California, on April 2–3. Some of the items discussed are highlighted below.

Postdoctoral Fellowships

Proposed genome-specific postdoctoral fellowships, modeled on the DOE Hollaender Fellowships, would involve similar competitive applications and joint DOE and institutional support. The fellowships would be tenable at any DOE-supported university or laboratory project of $150,000 or more. Short-term fellowships patterned on the European Molecular Biology Organisation (EMBO) program were also considered.

Reports

HGCC heard reports from several genome organizations, including the Joint Working Group on Mapping (p. 9) and those given below.

Ethical, Legal, and Social Issues. A report was given on the February Workshop of the Joint Working Group on Ethical, Legal, and Social Issues (ELSI) Related to Mapping and Sequencing the Human Genome. The role of ELSI is to anticipate problems and issues arising from expanded genetic information and from the use of human materials and samples. The following priority items were discussed: education of the public and professional groups, and confidentiality and use of test data. The working group may evaluate the cystic fibrosis experience as a model for handling information resulting from the genome project.

HUGO. The Human Genome Organisation (HUGO) reported that new funding for the administration and promotion of efforts to create working groups for each chromosome may contribute to international coordination of the genome effort. HUGO anticipates that expanding its membership will increase international communication and cooperation. DOE is actively involved with HUGO, and members of HGCC are among HUGO's current members. Charles R. Cantor, HGCC Chair, is Vice President of HUGO for the Americas region.

Sequencing Initiation

In a separate strategy session, HGCC discussed the possibility of starting to sequence DNA, particularly cDNA, earlier in the project than previously planned. There were two major factors in this discussion: (1) initiation of sequencing on a small scale because of the availability of more advanced sequencing technologies and (2) the realization that sequenced cDNA could be used as functional sequence tagged sites (STS) for mapping studies. Additionally, cDNA sequence identification would reveal functionally important DNA regions, since these DNAs are copies of mRNA. Establishment of chromosome-specific cDNA libraries and clones will be discussed during future meetings of HGCC, which is scheduled to meet June 19 in Bethesda, Maryland. ◇

Reported by Sylvia J. Spengler
Human Genome Center
Lawrence Berkeley Laboratory

Use Subscription/ Document Request Form on p. 20 to request copies of these publications:

- **HGCC minutes,**
- **5-Year Plan, and**
- **DOE Program Report.**

DOE-NIH 5-Year Plan, DOE Program Report Now Available from HGMIS

Documents Mailed to Subscribers

The following new publications were mailed in May to *Human Genome News* subscribers. These documents are available to anyone requesting copies.

Understanding Our Genetic Inheritance, The U.S. Human Genome Project: The First Five Years (FY 1991–1995), the 89-page joint DOE-NIH report, was published in April. The document states the scientific goals for the next 5 years, examines the current state of genome science, sets forth the complementary approaches of the two agencies, describes collaboration among U.S. and international groups, and presents budget projections for the project.

Human Genome 1989–90 Program Report, the 157-page DOE document published in March 1990, reports on the status of the OHER Human Genome Program and gives a brief background of the program, along with projected goals for the next 15 years. A research highlights section, narratives on major genome research efforts, and abstracts of work in progress are of special interest. Figures and captions provided by investigators give detailed information. ◇

DOE-NIH Subcommittee Establishes Joint Mapping Working Group

Working Group Promotes Redirection of Efforts to Mapping

At its December 1989 meeting, the Joint DOE-NIH Subcommittee on the Human Genome established the Joint Mapping Working Group to consider issues related to the project goals of genetic linkage mapping and physical mapping.

The working group's (see upper box for membership) first meeting was held in March in Salt Lake City to discuss the rate of progress in human genetic linkage mapping, an area discussed at the December meetings of the NIH Program Advisory Committee on the Human Genome (PACHG) and the Joint DOE-NIH Subcommittee.

Working group members David Botstein, David R. Cox, and Maynard Olson, along with a panel of invited consultants (see upper box), met to review the status of genetic mapping. They concluded that laboratory managers would agree to a formal contract redirecting a portion of their effort to mapping. Responsibility for mapping specific chromosomes is being undertaken.

In addition, the group plans to coordinate the identification of index markers on specific chromosomes. The near-term goal is to establish 300 markers located an average distance of 10 cM apart; increased focus will be on regions currently without markers. Further discussions on the status of each chromosome map will be held later this year.

The six-member Joint Mapping Working Group met the next day at Los Alamos National Laboratory (LANL), where discussions focused on developing an operational definition of sequence tagged sites (STS), on using STS as a mapping tool, and on making STS a common language among laboratories. The group agreed that sequencing a small segment of each marker will help to link the genetic map and the physical map. They noted that characterization of an STS must include proof of "single copiness"; the polymerase chain reaction (PCR) product of STS must be a single band in Southern hybridizations.

In addition, the group considered certain informatics needs: characteristics of the laboratory notebook database for the next 10 years and procedures for handling notebook material.

Joint Mapping Working Group

David Botstein	Genentech, Inc.
Anthony V. Carrano	Lawrence Livermore National Laboratory
C. Thomas Caskey	Baylor College of Medicine
David R. Cox	University of California at San Francisco School of Medicine
Robert K. Moyzis	Los Alamos National Laboratory
Maynard V. Olson	Washington University School of Medicine

Invited Consultants

Norman Arnheim	University of Southern California at Los Angeles
Helen R. Donis-Keller	Washington University School of Medicine
James Gusella	Massachusetts General Hospital
Leroy E. Hood	California Institute of Technology
Eric Lander	Whitehead Institute for Biomedical Research
Jeffrey C. Murray	University of Iowa Hospital
Raymond L. White	University of Utah Medical School

Reports and recommendations of both working group meetings were to be presented to the NIH PACHG and the Joint DOE-NIH Subcommittee at their meetings in June. Recommendations by these committees will be discussed in the next issue of *Human Genome News*. ◇

Reported by Mark Guyer
Assistant Director for Program Coordination
NCHGR

NCHGR Invites Grant Applications for Support of Short-Term Advanced Courses

The National Center for Human Genome Research (NCHGR) invites applications from academic and research institutions for grants to support short-term advanced courses to enhance the skills of individuals interested in pursuing laboratory or scholarly research in areas related to the Human Genome Project. These courses would emphasize new laboratory techniques in genome analysis; informatics; principles of genome analysis for nonbiology scientists and scholars in humanities, social science, or law; and principles and methods of studying relevant social, ethical, and legal issues.

The goal of courses, typically 1 to 2 weeks long and offered annually, is to improve the level of cross-disciplinary training so that scientists and other professionals can participate more fully in the Human Genome Project and can use information and technology from the project in other areas.

Application kits are available from most institutional business offices and from the Office of Grants Inquiries; Division of Research Grants; NIH; Westwood Building, Room 449; 5333 Westbard Avenue; Bethesda, MD 20892; (301) 496-7441. **Deadline for applications is August 24.**

Potential applicants are encouraged to discuss plans and objectives of their proposed courses with NCHGR staff before applying. For more information and a copy of the Request for Application (#HG-90-01), contact: Bettie J. Graham, NCHGR, (301) 496-7531; Fax: (301) 480-2770; Bitnet: "b2g@nihcu "; Internet: "b2g@cu.nih.gov" . ◇

Genome Project News

Joint Informatics Task Force Identifies Issues

The first meeting of the NIH-DOE Joint Informatics Task Force (JITF) was held in Bethesda, Maryland, March 7–8. JITF is a task force set up to offer guidance and to help the agencies coordinate work in genome informatics. For a complete description of the goals and purposes of JITF and a list of task force members, see *Human Genome News*, Vol. 2, No. 1 (May 1990), pp. 10–11.

The task force, which meets with members of its Liaison Group (see box), identified several major issues at the March meeting:

1. Data Requirements.

Participants discussed the kinds of data the Human Genome Project may produce and how the data should be stored and maintained.

2. Connectivity and Infrastructure.

Internet connections are a minimum requirement for genome centers and would be of great benefit to researchers everywhere. To encourage people to use computer technology, the network should provide interesting and useful resources; funding and practical advice for making use of informatics should also be supplied.

3. Training.

Because genome informatics is based on the interaction of biological scientists with experts in statistics, computer science, and engineering, the interdisciplinary training of scientists is needed.

4. Long-Term Needs and Goals.

The success or failure of past large-scale data projects and the use of past experiences in planning genome project informatics were discussed, and some desirable information technologies not yet available were identified.

Based on the issues raised, four working groups were created to address specific issues. Each working group and its chair are listed in the box.

The task force recommended the following issues for consideration:

- publication, in some form, of all future databases funded by agencies involved in genome research;
- inclusion of an explicit syntax (computer-parsible) for the data collection described in grant proposals for database development, as well as public availability of detailed documentation of this syntax;
- establishment of mechanisms and funds for collection of small curator-based databases; and
- more detailed specifications by funding agencies regarding genome informatics issues in instructions to both applicants and reviewers.

Initial reports from the working groups are expected by August, and the next meeting of JITF will be held in October.

Periodic progress reports from the task force will be disseminated so that the scientific community can respond in detail. Input should be directed to JITF members or liaisons or to the working group chairs as appropriate. See box for E-mail addresses. ◊

Submitted by Gregory Hamm
Molecular Biology Computing Laboratory
Waksman Institute/CABM
Rutgers, The State University of New Jersey

Liaison Members

James Cassatt	National Institute of General Medical Sciences
Diane Hinton	Howard Hughes Medical Institute
Elke Jordan	National Center for Human Genome Research
David Lipman	National Library of Medicine
Robert Robbins	National Science Foundation
Keith Russell	National Agriculture Library
Jeffrey Schmaltz	DOE

Working Groups

Data Requirements	Connectivity and Infrastructure	Training	Long-Term Needs and Goals
Frank Olken*	Gregory Hamm*	Sylvia Spengler*	Eric Lander*
Elbert Branscomb	Elbert Branscomb	John Devereux	David Botstein
Peter Pearson	Frank Olken	Diane Hinton	Nathan Goodman
Thomas Marr	Robert Robbins	Mark Pearson	Gregory Hamm
Michael Waterman			Thomas Marr

*Chair

Electronic mail can be sent to JITF members and liaisons using one of these addresses:

Internet "jitf@mbcl.rutgers.edu"; Bitnet: "jitf@mbcl"; Bitnet: "jitf@biovax".

Workshop on Chromosome 3

Participants Generate Consensus Maps

Laboratory representatives described their research and generated consensus maps of chromosome 3 at a meeting sponsored by NIH and the RGK Foundation. Held in San Antonio at the University of Texas Health Science Center (UTHSC) on February 16–17, the chromosome-3 meeting attracted 42 participants from laboratories in Canada, England, Japan, the Netherlands, Sweden, and the United States. Highlights of some of the papers are included below.

Physical Mapping

David Smith (Wayne State University) disclosed the isolation of a large number of chromosome-3-specific cosmids and the identification and mapping of several "cluster" cosmids containing multiple rare restriction sites. Harry Drabkin [Eleanor Roosevelt Institute for Cancer Research (ERICR)] reported on a number of markers and the establishment of a radiation hybrid panel. Lakshmi Atchison (Fox Chase Cancer Center) described several projects, including rare-site linking libraries, isolation of markers, mapping by in situ hybridization, and construction of a radiation hybrid panel. Pamela Rabbitts [Medical Research Council (MRC) Cambridge, U.K.] has mapped several markers by in situ hybridization and firmly established reference points; she has also determined that ErbAβ is the same locus as ErbA2.

Michael Lerman [National Cancer Institute (NCI), Frederick Cancer Research Facility (FCRF)] reported on 2000 single-copy lambda clones isolated from flow-sorted libraries and mapped to regions of chromosome 3; of these clones, 53 were found to detect useful restriction fragment length polymorphisms (RFLPs). Susan Naylor (UTHSC) described a panel of radiation hybrids constructed from a *neo*-marked chromosome 3 and the identification of polymerase chain reaction (PCR) primers for 40 transcribed genes on chromosome 3. Ben Carritt (MRC, London) discussed origination of the D3F15S2 locus and its related sequence on chromosome 1. Robert Gemmill (ERICR) presented a detailed pulsed-field gel map for the region p14 to p21.1; in addition, his group has identified two probes that flank the t(3;7) breakpoint. York Miller (ERICR) reported the cloning of a cDNA for aminoacylase 1 (ACY 1 at 3p21.1–3p24.2.); this widely transcribed enzyme is not detected in 20% of small-cell lung cancers.

Linkage Studies

New data on von Hippel-Lindau (VHL) syndrome linkage, reported by Bernd Seizinger (Massachusetts General Hospital), included markers developed with David Smith and Harry Drabkin; the data suggest that VHL maps in a telomeric position to RAF1, and that a single gene codes for VHL. Confirmation of the VHL-RAF1 linkage was reported by Jeff Vance (Duke University) and Eamonn Maher (Cambridge University).

Bert Zbar (NCI, FCRF) presented linkage data in VHL families and a revised linkage map of markers on 3p. Phyllis McAlpine (University of Manitoba) detailed linkage data from a family containing an inversion of chromosome 3 [inv(3)(p25q21)]. She also reported on the cloning of chromosome-3 segments in yeast artificial chromosomes (YACs).

Vincent Stanton, Hiroyuki Aburatani, and David Housman (Massachusetts Institute of Technology) reported on the conversion of several RFLPs to PCR assays, which they used in meiotic mapping studies with Norman Arnheim (University of Southern California). Takaaki Sato (Cancer Institute, Tokyo) reported a chromosome-3-specific library from which 4000 cosmid clones have been isolated. Twenty new RFLPs are being placed on the chromosome-3 linkage map.

Comparative Mapping and Informatics

Peter Lalley (Wayne State University) is using comparative mapping to investigate the evolutionarily conserved human chromosome-3-specific sequences found on mouse chromosomes 9, 6, and 16.

A database system for maintaining genetic information for chromosome 3 was discussed by Tim Bishop (Imperial Cancer Research Fund, Leeds, U.K.). The database structure is designed to incorporate the name and location of each locus and information on the sequence, polymorphism, and primer sequences required. ◇

Reported by Susan Naylor
Department of Cellular and Structural Biology
University of Texas Health Science Center
at San Antonio

Chromosome-3 meeting contact:

Susan Naylor
(512) 567-3842

Genetic and physical maps determined from this workshop will be included as part of the HGM 10.5 report. The next workshop on chromosome 3 will be held in about a year.

Meeting Reports

International Workshop on Human Chromosome 21

Investigators Agree To Consolidate Map Data

Some 35 investigators at the international Workshop on Human Chromosome 21 agreed to consolidate their data into a unified map of the chromosome. Held April 2–3 in Bethesda, Maryland, the workshop was organized by David Cox (University of California) to establish the current status of both the physical and the genetic maps of chromosome 21.

The meeting consisted primarily of round-table discussions in which investigators presented their published and unpublished data. Many workshop participants felt that the chromosome-21 research community set a standard of cooperation likely to be emulated in many large-scale mapping efforts.

The meeting was successful on many levels; perhaps the most encouraging measure of success was the cooperation of participants in consolidating their data into a unified map of the chromosome, resulting in the definition of composite physical and genetic maps. Participants chose 23 reference markers, whose order on the chromosome is established, to define the physical map of chromosome 21. The nine markers serving as "anchor markers" for the genetic linkage map of chromosome 21 provide a common language and should facilitate assignment of additional markers to genetic and physical maps of chromosome 21.

Another achievement of the meeting was the establishment of the Chromosome-21 Joint YAC Effort to consolidate activities of the chromosome-21 research community in screening yeast artificial chromosome (YAC) library resources. David Patterson (Eleanor Roosevelt Institute for Cancer Research), whose laboratory is receiving a copy of the human YAC library from the Center for Human Genetics at Washington University, will screen the library on behalf of the community; screening is expected to minimize duplication of effort and to make the best use of resources. A number of laboratories will make their chromosome-21-specific YAC clones available to this international project for screening.

A YAC newsletter—*FAX on the YACs*—will be circulated to active participants in the Chromosome-21 Joint YAC Effort. It will coordinate production of sequence tagged sites (STS) for screening purposes and will generally report to the community the identity and integrity of the YACs being isolated. All isolated YACs will be freely available to participants in the joint YAC effort. ◇

Reported by Sue Rider and David Cox
Department of Biochemistry
University of California at San Francisco
School of Medicine

For information on the Joint YAC Effort, contact:

Sue Rider
Fax: (415) 476-8001
Rudy Tanzi
Fax: (617) 726-5735

Mathematical Approaches to DNA

"Mathematical Approaches to DNA," the first meeting sponsored by the National Science Foundation (NSF) Program in Mathematics and Molecular Biology, was held on January 24–28 in Santa Fe, New Mexico. It attracted over 120 participants, including mathematicians, computer scientists, biologists, chemists, and physicists. Topics included consideration of the mechanics of DNA motion in electrophoresis gels and algorithms for macromolecule structural sequence matching, gene identification, and identification of structural motifs predictive of function.

Dynamic programming algorithms with performances comparable to or faster than FASTA and FASTP sequence alignment software programs were discussed by Patrick Powell (University of Minnesota) and William Pearson (University of Virginia). A method using multiple alignments as opposed to single sequences, developed by Robert Jones (Thinking Machines, Inc.), permits detection of weak-sequence relationships and identification of biologically relevant sites and domains.

Work from the Theoretical Biology and Biophysics Group of Los Alamos National Laboratory (LANL) focused on (1) automatic identification of pattern characteristics for gene identification (Gwennaele Fichant), (2) analysis of phylogenetic relations in repeated DNA sequences families (Yves Quentin), and (3) mathematical problems

(see Mathematics, p. 17)

Cold Spring Harbor Meeting on Genome Mapping and Sequencing

Applications of New Technologies May Reduce Project Cost and Labor

Progress in several areas of genome mapping and sequencing suggests that technology may reduce the cost and labor needed to achieve genome project objectives. This theme was repeated several times during the recent 1990 Cold Spring Harbor Laboratory meeting on Genome Mapping and Sequencing. The organizers – Charles Cantor, Maynard Olson, and Richard Roberts – planned an excellent and timely program for the May 2–6 meeting in Cold Spring Harbor, New York.

During last year's meeting, several new technologies were suggested and discussed, but this year scientists report that quite a few of them are actually being applied.

Round worm mapping. The project showing most advancement is the mapping (and soon sequencing) of the round worm *Caenorhabditis elegans* genome, which may serve as a technological pilot project for the much larger human genome. Groups directed by John Sulston (MRC Laboratory for Molecular Biology, Cambridge, U.K.) and Robert Waterston (Washington University, Missouri) have nearly completed the physical map of *C. elegans*. Equally important, they have helped establish the utility of yeast artificial chromosomes (YACs) for large-scale mapping.

YACs. YACs, frequently mentioned at the meeting, dominated the session on large DNA cloning methods. Inserted DNA was reported to maintain its integrity in YACs, and the use of overlapping YACs is gaining importance in conjunction with cosmid mapping.

Automation. The topic of automation also recurred during the meeting. A new session on automation highlighted methods for mapping, sequencing, and data handling, and discussions in other sections indicated that automation is now a concern of many investigators.

PCR. Polymerase chain reaction (PCR) continues to generate advances in sequencing, as it has in other areas of molecular biology. A big advantage to using PCR is that it may permit both DNA strands to be sequenced simultaneously.

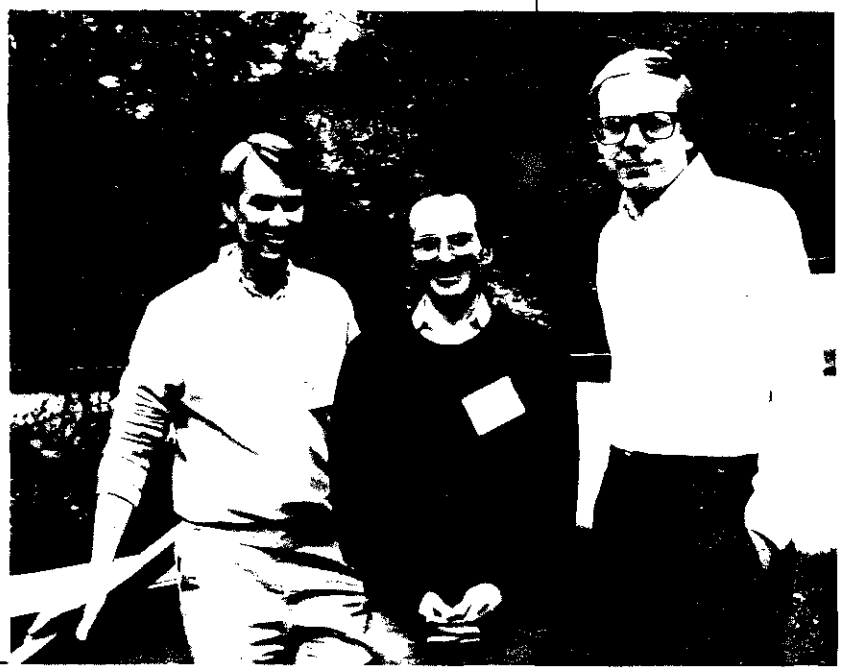
Oligonucleotide primers. Libraries of oligonucleotide primers 9 or 10 nucleotides in length could greatly reduce the cost and effort of high-volume sequencing. Currently, the feasibility of this strategy is being tested.

Separation by magnetic fields. Magnetic beads, in combination with the biotin-streptavidin system, are being used to separate DNA strands after PCR, eliminating the need for template purification before solid-phase sequencing of genomic and plasmid DNA. The use of magnetic fields can be automated more readily than centrifugal force separation, because automating centrifuge loading would require very sophisticated robotics.

Cleavage strategies. Combinations of methylases, restriction enzymes, and DNA binding proteins can be used to cut genomic DNA at very specific sites. This new method has been used to make single and double cleavages in the *Escherichia coli* genome and a single cleavage in an entire yeast genome. ◇

Reported by Daniel R. Schechter
Science Writer
Cold Spring Harbor Laboratory

Round Worm Genome Mapping May Serve as Pilot Project



Pictured (l. to r.) are **Richard Roberts, Charles Cantor and Maynard Olson – organizers of the Cold Spring Harbor Laboratory meeting on Genome Mapping and Sequencing.**

Meeting Reports

Electrophoresis, Supercomputing, and the Human Genome

The First International Conference on Electrophoresis, Supercomputing, and the Human Genome was held April 10–13 in Tallahassee, Florida. Hosted by the Supercomputer Computations Research Institute (SCRI) of Florida State University, the meeting attracted about 110 international participants, including representatives from France, Japan, Sweden, the United Kingdom, the United States, the U.S.S.R., and Yugoslavia. The interdisciplinary conference was planned to foster the exchange of ideas and information regarding improvement of technology and involved computational experts, experimentalists, and technologists.

Since the start of the genome initiative, the necessary involvement of scientists with widely divergent backgrounds has been evident. The correct handling, analysis interpretation, and dissemination of data and information and the control and data gathering of automated processes are areas where computer science is directly involved. Possibly not so obvious, but of increasing importance, is the use of computers to model complex phenomena such as pulsed-field gel (PFG) electrophoretic parameters during DNA fragment separation. The desire to understand provided common ground for scientists from all disciplines at this meeting. Program highlights are given below.

Topics Presented

- Automated contig mapping system
- Supercomputer modeling of chromosomes
- Computer modeling of PFG electrophoretic parameters
- Electrophoresis system modifications
- Computer analysis of protein structure
- Modifying DNA to facilitate separation of larger strands
- Sequencing by hybridization
- Discussions on international genome efforts

Computing Applications

George I. Bell [Los Alamos National Laboratory (LANL)] opened the meeting with an excellent overview of the involvement of computers in the Human Genome Project, emphasizing computerization of overlap probabilities for contig assembly, sequence similarity comparisons, and identification of intron-exon boundaries.

Tom Duke (TCM/Cavendish Laboratory, U.K.), Bengt Norden (Chalmers Institute of Technology, Sweden), Jean Louis Viovy (CNRS Laboratorie de Physiochimie Theoretique, France), J. M. Duestch (University of California at Santa Cruz), and Y. P. Papov (U.S.S.R. Academy of Science) made presentations on theoretical modeling and computer simulations of the serpentine movement of DNA during PFG electrophoresis. Nancy Stellwagen (University of Iowa) described her experimental results on orientation effects of electric fields on agarose gels. Electric birefringence studies of gels and preliminary studies with DNA indicate that DNA migration can be influenced by orientation and reorientation of the gel matrix by the electric field.

Anthony V. Carrano (Lawrence Livermore National Laboratory) described automation and computerization currently being used to construct physical maps. A unique set of computer programs, a database management system, and a workstation computer network complement automatic contig mapping using restriction enzymes. Over 4000 cosmids have been assembled into about 400 contigs for chromosome 19.

In the realm of supercomputers, William Shoaff (Florida Institute of Technology) presented his work on building a supercomputer model of a human chromosome. Parallel computer architecture was not neglected at this conference; Robert Jones (Thinking Machines, Inc.) reported his work on multiple sequence alignment using a massively parallel machine.

Laboratory Applications

Many interesting papers were presented on gel electrophoresis. Using a modified electrophoresis system with high voltages and efficient cooling, Eric Fairfield (LANL) achieved separation by agarose gel electrophoresis of fragments up to 4 kb in length in very short running times.

An experimental highlight presented by Levy Ulanovsky (Harvard University Biolabs) described a new technique for binding a bulky protein (streptavidin) to the end of single-stranded DNA; modifying DNA in this way facilitates separation of larger single strands on polyacrylamide gels. Applied to DNA sequencing, this technology promises to help raise the 250- to 300-bp upper limit on resolution caused by the inability to resolve large single-stranded fragments on polyacrylamide gels.

Y. P. Lysov (Engelhardt Institute of Molecular Biology, U.S.S.R.) presented a general overview of random oligo hybridization sequencing. R. Drmanac (Imperial Cancer miniaturized "sequencing chip" for rapid

(See Electrophoresis, p. 16)

Conference proceedings (ISBN: 981-02-0273-3) will be published by World Scientific Publishing Co., Ltd., (201) 837-8858, and will be available in December.

AI and Molecular Biology Symposium

Results Presented Suggest Importance of Combining Multiple Reasoning Techniques

The Stanford University 1990 Spring Symposium Series, sponsored by the American Association for Artificial Intelligence (AAAI), was held March 27–29. For the first time, the series included an Artificial Intelligence (AI) and Molecular Biology symposium, which proved to be an excellent forum for exchange of ideas among AI scientists and engineers who are concentrating on research applications in molecular biology.

Abstracts were submitted from more than 160 researchers in Australia, Canada, England, Greece, Italy, Japan, Scotland, and the United States; the large number of researchers and the breadth of investigations using AI techniques indicate the rapid growth in this interdisciplinary field within the last 5 years.

The symposium featured 30 short technical presentations, grouped into the 8 sessions shown in the table below.

Technical Presentation Sessions

- **Protein Structure – Expert Systems and Pattern Recognition**
- **Protein Structure II – Machine Learning**
- **Modeling Biological Processes**
- **Minimal Length Encoding and Molecular Biology**
- **DNA Sequence Analysis I – Problem Solving**
- **DNA Sequence Analysis II – Machine Learning and Neural Nets**
- **Gene Mapping**
- **Data Analysis Aids**

The session on Money, Careers, and Resources provided an opportunity for Peter Clepper (National Library of Medicine) and Robert Futrelle (Northeastern University) to elaborate on funding mechanisms and resources at NIH and the National Science Foundation (NSF), respectively. The concluding session on Senior Perspectives prepared by David Searls (Unisys), Doug Brutlag (Stanford University School of Medicine), Peter Friedland (NASA Ames Research Center), and Joshua Lederberg (Rockefeller University) explored the notion of DNA as a language and the future roles of AI in molecular biology research.

In a final discussion led by program chairman Lawrence Hunter (National Library of Medicine), participants noted that molecular biology, with its "constrained complexity," represents a logical application area for AI and related methods and algorithms because problems posed by molecular biology research are at the appropriate level of complexity for current AI techniques. Established standards are available to assess the performance of new or reenacted methodologies, and funds seem to be sufficient to support an effective overall effort based on peer review. Additionally, collaborations appear to be readily possible because personnel from both disciplines are represented at most research centers and universities. ◇

Submitted by Reinhold C. Mann, Head
Intelligent Systems Section
Engineering Physics and Mathematics Division
ORNL

A listing of the contents of a database of over 100 researchers and funding agents in AI and Molecular Biology is available. The purpose of this database is to facilitate communication with researchers in biology and computer science. Via anonymous ftp on the Internet, the database may be downloaded from the directory "/pub/aimb-db" on host "lhc.nlm.nih.gov". Internet address: 130.14.1.128.

To request files via e–mail, send a message to:
Lawrence Hunter
National Library of Medicine
Internet: "hunter@nlm.nih.gov"
Bitnet: "hunter%nlm.nih.gov@nihcu"
(301) 496-9300
Fax: (301) 496-0673

NAS Computing and Molecular Biology Meeting

Computational molecular biology may receive an influx of fresh talent, based on the initial reaction of participants at a recent Computer Science and Technology Board (CSTB) workshop. "Computing and Molecular Biology: Mapping and Interpreting Biological Information" was held April 30–May 1 at the National Academy of Sciences (NAS) in Washington, D.C.

Sponsored by CSTB, the workshop highlighted the important and challenging information analysis problems that exist in molecular biology. In keeping with workshop objectives, about two-thirds of the participants were computer scientists, half of whom were in early stages of their research careers. While many came with only a cursory knowledge of molecular biology, they were uniformly enthusiastic about applying their expertise to this field. Molecular biologists and representatives from funding agencies contributed to an informative dialogue.

(see NAS, p. 17)

Computer Science Participants Enthusiastic About Applying Expertise

For more information on NAS meeting, contact:

Damian Saccocio
National Research Council
CSTB/HA 560
2101 Constitution Avenue, NW
Washington, DC 20418

Resources

GDB Forum, a quarterly newsletter, will keep members of the scientific and medical communities informed about the development of the Genome Data Base.

To receive this publication, contact:

Director
Genome Data Base
William H. Welch
Medical Library
1830 E. Monument
Street, 3rd Floor
Baltimore, MD 21205
(301) 955-9705
Fax: (301) 955-0054

Johns Hopkins University Launches Genome Data Base

Johns Hopkins University has launched the Genome Data Base (GDB), a networked resource designed to support worldwide efforts to map and sequence the human genome. GDB will collect, organize, store, and distribute genetic mapping information. The database also will serve as a repository for genetic disease information applicable to patient care. First release of GDB is scheduled for September.

With initial support from the Howard Hughes Medical Institute (HHMI), and backed by the resources of the Johns Hopkins School of Medicine and the Laboratory for Applied Research in Academic Information at the William H. Welch Medical Library, GDB builds on decades of genetic data generation and collection. Johns Hopkins is also home to the *On-line Mendelian Inheritance in Man* (OMIM), developed by geneticist Victor A. McKusick. OMIM will be linked to GDB.

Located at the Welch Library on Johns Hopkins' East Baltimore campus, GDB will collaborate with other publicly and privately managed databases. Through an agreement with the Imperial Cancer Research Fund in London, for example, GDB will serve as the official database for Human Gene Mapping Workshops 10.5 and 11.

GDB scientific director is Peter L. Pearson, and Richard E. Lucier, director of the Welch Laboratory, is responsible for database development and service aspects. GDB editors, chairs, and cochairs of the Human Gene Mapping Workshops will maintain database quality and currency. ◊

Reported by Kim Goad
Publications Coordinator
The Johns Hopkins Medical Institutions

Human Genome Abstracts Begins Publication

Human Genome Abstracts: Basic Research and Clinical Applications (ISSN: 1045-4470) began bimonthly publication in February. Some 600 abstracts, mainly from published literature relating to the Human Genome Project, with subject and author indexes, are selected from over 5000 sources for each issue.

Chairing the publication's advisory board is Cassandra L. Smith, affiliated with the Human Genome Center at Lawrence Berkeley Laboratory and with the Department of Molecular and Cell Biology at the University of California at Berkeley.

The print version is available from Cambridge Scientific Abstracts, (301) 961-6750. The online version is available in Dialog File 76, Life Sciences Collection; Dialog Information Services, Inc.; (800) 334-2564. ◊

Electrophoresis *(from p. 14)*

Research Fund, London) reported on developing a miniaturized "sequencing chip" for rapid hybridization and readout using image processing and on experimental work using 6- to 8-mer probes of a human fetal brain cDNA library on nylon membranes. Both speakers emphasized the importance of computers for interpretation and quantification of sequencing-by-hybridization (SBH) results.

Other presentations ranged from in-gel DNA reassociation by Michio Oishi (University of Tokyo) and computer analysis of protein structure by Y. V. Sergeev (U.S.S.R. Academy of Science) to an informative overview of GenBank® by Paul Gilna (LANL).

Overviews and Conclusions

Evening sessions included informative overviews of their respective country's genome effort by Alexander A. Bayev (U.S.S.R. Academy of Science), Charles R. Cantor (Lawrence Berkeley Laboratory) and Charles DeLisi (Boston University), and Nobuyoshi Shimizu (Keio University School of Medicine, Japan).

Conclusions from the conference include the following:

- The DOE-NIH 5-Year Plan is achievable.
- Improvements in technologies and instrumentation are necessary.
- Computers will become more important to the project because they are needed for a number of applications, including:
  - genome workstations for accessing a series of databases and for manipulating and updating data to accomplish gene sequence analysis;
  - supercomputers for simulating protein structure and function, for molecular recognition, and for determining macromolecular folding; and
  - automated programmable machines for the sequencing process.
- Animal models are needed to correlate with human genome data. ◊

Reported by Richard J. Douthart
Life Sciences Center
Battelle Pacific Northwest Laboratories
and
Hwa A. Lim
SCRI
Florida State University

Databases Organize Information Resources

Two databases containing information about other databases related to genome research and genome projects are available to researchers and other persons interested in the Human Genome Project. The Listing of Molecular Biology (LiMB) at Los Alamos National Laboratory (LANL) describes the means of accessing and the contents of many foreign and domestic databases related to molecular biology. The Directory of Biotechnology Information Resources (DBIR™) of the National Library of Medicine (NLM) contains information on a wide range of resources related to biotechnology.

LiMB Database Available on Disk in Searchable Format

Entries in LiMB provide information such as database history, names and addresses of contact persons, ways to obtain the data, obsolete or incorrect names, types of hardware and software used, limitations on access, and frequency of updates. A dictionary defining the contents of each entry precedes the body of data and makes searching easy.

Currently, the database has 80 entries and is updated about every 6 months. Release 1.2 is now available, and the 2.0 updated version will be offered later this year. The LiMB database is available in searchable format on floppy disk. See side column for contact information.

DBIR™ Database Accessible Online

DBIR™, an online multicomponent biotechnology database, has over 1400 records that are added to and updated monthly; listed organizations and entities are contacted yearly for changes. Included is information on resources such as (1) other computerized databases and their distributors, data networks, electronic bulletin boards, and other computerized resources for collecting and disseminating biotechnology data; (2) publications that focus on biotechnology issues, including directories, serials, monographs, journals, reviews, and compilations; (3) committees addressing issues of nomenclature and standards in biotechnology and molecular biology; (4) biological culture collections and specimen banks; and (5) biotechnology centers and other organizations that foster development. Data is assembled by the American Type Culture Collection. ◇

Submitted by Laura N. Yust
HGMIS, ORNL

LiMB contacts:
Gifford Keen (LiMB Administrator) and Christian Burks
Theoretical Biology and Biophysics Group, LANL
(505) 667-9455
Internet: "limb@lanl.gov"

The DBIR™ database can be accessed via the NLM TOXNET® in NLM's ELHILL system.

DBIR™ contact: James J. Ferguson (301) 496-6531

NAS *(from p. 15)*

The workshop was cochaired by Robert Langridge (University of California at San Francisco) and Eric Lander (Whitehead Institute for Biomedical Research). Major areas addressed included sequence analysis, information storage and retrieval, and protein structure prediction. Specific challenges presented to the participants included:

- finding genes within a DNA sequence;
- deducing structure and function from protein sequence;
- reconstructing evolutionary trees from sequence probability;
- processing experimental data; and
- facilitating the storage, access, and update of distributed data.

To continue the cross-disciplinary dialogue begun at the workshop, short-term courses in molecular biology and computer science, problem-specific workshops, standard datasets, joint research projects, and a mailing list of relevant researchers are being considered. A summary of the workshop will be available later this year. ◇

Reported by Damian Saccocio
Computer Science and Technology Board, NRC

Mathematics *(from p. 12)*

associated with construction of ordered clone maps and incomplete restriction maps of human chromosomes (David Torney).

The program is funded by a grant from the Division of Mathematical Sciences and the Division of Biological, Behavioral, and Social Sciences of NSF. IBM provided additional funding for this meeting. (For more details, see "Mathematics Untwists the Double Helix," *Science*, Feb. 23, 1990, pp. 913–915.)

Title of the meeting planned for March 1991 will be "The Genome: Mathematical Analysis from DNA to Protein." ◇

Reported by Sylvia J. Spengler, Associate Director
Program in Mathematics and Molecular Biology
University of California at Berkeley

Mathematics meeting contact: Sylvia J. Spengler, (415) 643-7799, Internet: "sylviaj@violet.berkeley.edu"

Calendar of Genome Events*

July	9–11	**International Meeting: Bioinformatics, Integration of Organismic and Molecular Data Bases, and Use of Expert Systems in Biology**; Fairfax, VA [*H. Morowitz, (703) 323-2262, Fax: (703) 764-4725*]
	18–21	**Genetics Societies of America and Canada Joint Meeting**; San Francisco [*J. Francese, (301) 571-1825, Fax: (301) 530-7079*]
	22–27	**CIOMS 24th Conference—Genetics, Ethics, and Human Values: Human Genome Mapping, Genetic Screening, and Genetic Therapy**; Tokyo
	23	**5th International Conference on Scanning Tunneling Microscopy/ Spectroscopy** and **1st International Conference on Nanometer-Scale Science and Technology**; Baltimore [*J. Murday, (202) 767-3026*]
	25–27	**9th Summer Symposium in Molecular Biology**; University Park, PA [*B. LaPorte, (814) 863-3696*]
August	2–4	**International Workshop on Chromosome 19**; Charleston, SC
	27–30	**"Symposium on Mapping and Sequencing"** at the **1990 National American Chemical Society Meeting—Analytical Division**; Washington, DC [*L. Smith, (608) 263-2594*]
September	6–11	**International Workshop on Human Gene Mapping (HGM 10.5)**; Oxford, U.K.
	13–16	**Converging Approaches in Computational Biology**; Rensselaerville, NY; application deadline: July 30 [*C. Keith, (518) 442-4327, Fax: (518) 442-4767*]
	18	**DOE Human Genome Coordinating Committee**; Livermore, CA
	30–Oct. 3	**Genome Sequencing Conference II**; Hilton Head SC; abstract deadline: July 15 [*S. Wallace, (301) 480-0634, Fax: 301) 480-8588*]
October	1–3	**First International Conference on DNA Fingerprinting**; Berne, Switzerland [*G. Dolf, Fax: (Int.) 031-24-7021*]
	16–20	**American Society of Human Genetics Annual Meeting**; Cincinnati, [*J. Francese, (301) 571-1825, Fax: (301) 530-7079*]
	22–24	**Human Genome II: An International Conference on the Status and Future of Human Genome Research**; San Diego; abstract deadline: August 15 [*Scherago Assoc., Inc., (212) 730-1050, Fax: (212) 382-1921*]
November	12–14	**2nd Workshop on International Cooperation for the Human Genome Project: Ethics**; Monte Picayo, Spain
	14–16	**Conference on the Impact of Biotechnology on Health Care**; Barcelona, Spain [*P. Moon, Oxford, U.K., (Int.) 44-865-512242, Fax: (Int.) 44-865-310981*]
December	3	**NIH Program Advisory Committee on the Human Genome**; Bethesda, MD [C. Mohan, (301) 496-0844]
	4	**DOE-NIH Subcommittee on the Human Genome**; Bethesda, MD
	4	**HGCC DOE Human Genome Coordinating Committee**; Bethesda, MD

*Attendance at meetings listed without contact information is by invitation only.

Calendar of Genome Events

January 1991	8–11	**"Biotechnology Computing Minitrack"** at the **Hawaii International Conference on System Sciences–24**; Kailua-Kona, HI; abstract deadline: June 6, 1990 [*L. Hunter, (301) 496-9300, Fax: (301) 496-0673*]
	27–Feb. 1	**Bio/Technology Magazine Winter Symposium — Advances in Gene Technology: The Molecular Biology of Human Genetic Disease;** Miami Beach [*The Miami Bio/Technology Winter Symposia, (800) 642-4363, Fax: (305) 324-5665*]
March 1991	19–21	**International Electrophoresis Society Meeting**; Washington, DC; abstract deadline: September 1, 1990 [*Janet Cunningham, (301) 898-3772, Fax: (301) 898-5596*]

*Attendence at meetings listed without contact information is by invitation only.

Training Calendar: Workshops and Coursework

July	9–13	**Recombinant DNA Techniques**; Germantown, MD (also offered July 30–Aug. 3 and Sept. 17–21) [*BRL Life Technologies, Inc., (800) 828-6686*]
	19	**DNA Amplification by PCR**; San Diego (also offered Aug. 20 and 28, Sept. 11, and Dec. 11 at other places) [*S. Chance, (515) 232-8306*]
	23–27	**Research Conference on Molecular Genetics**; Newport, RI [*A. Cruickshank, (401) 783-4011/3372*]
	25–Aug. 10	**"cDNA Cloning & Gene Expression," Carolina Workshops on Recombinant DNA Technology**, Chapel Hill, NC [*S. Kelly, (919) 966-1730*]
	31–Aug. 3	**Basic Cloning Techniques**; Northfield, MN (also offered Aug. 7–10) [*D. Betsch, (515) 232-8306*)]
August	12–17	**"Techniques: Physical Manipulation of the Human Genome"** and **"Human Genetics"** at the **Cellular and Molecular Genetics Research Conference**; Copper Mountain, CO [*M. Marsh, (301) 530-7093*]
	13–18	**Library Preparation and Analysis**; Miami [*D. Betsch, (515) 232-8306*]
	14–17	**Hybridization Analysis**; St. Louis (also offered Sept. 11–14 and 18–21) [*D. Betsch, (515) 232-8306*]
	20–26	**Workshop on cDNA Libraries**; Germantown, MD (also offered Sept. 24–30) [*BRL Life Technologies, Inc., (800) 828-6686*]
September	10–14	**Transfection Techniques**; Germantown, MD [*see contact Aug. 20–26*]

Acronym List

Acronyms listed were chosen because they were either used in the text or are relevant to the human genome research community. Listed in parentheses after an organization is the branch of government or the organization to which it is responsible.

*Denotes U.S. Department of Energy organizations.
†Denotes U.S. Department of Health and Human Services organizations.

AAAI	American Association of Artificial Intelligence
AI	Artificial intelligence
CCD	Charged-coupled display
cDNA	Complementary DNA
CSTB	Computer Science and Technology Board (NRC)
DBIR†	Directory of Biotechnology Information Resources (NLM)
DHHS	Department of Health and Human Services (U.S.)
DNA	Deoxyribonucleic acid
DOE	Department of Energy (U.S.)
ELSI*†	Working Group on the Ethical, Legal, and Social Issues
EMBO	European Molecular Biology Organisation
FCRF†	Frederick Cancer Research Facility (NCI)
GDB	Genome Data Base (HHMI, Johns Hopkins University)
HGCC*	Human Genome Coordinating Committee
HGMIS*	Human Genome Management Information System (ORNL)
HHMI	Howard Hughes Medical Institute
HUGO	Human Genome Organisation [international]
ICRF	Imperial Cancer Research Fund (U.K.)
JITF*†	Joint Informatics Task Force
LANL*	Los Alamos National Laboratory, Los Alamos, N.M.
LBL*	Lawrence Berkeley Laboratory, Berkeley, Calif.
LIMB	Listing of Molecular Biology (LANL)
LLNL*	Lawrence Livermore National Laboratory, Livermore, Calif.
MRC	Medical Research Council, U.K.
mRNA	Messenger ribonucleic acid
NAS	National Academy of Sciences (U.S.)
NCHGR†	National Center for Human Genome Research (NIH)
NCI†	National Cancer Institute (NIH)
NIH†	National Institutes of Health
NLM†	National Library of Medicine
NRC	National Research Council (NAS)
NSF	National Science Foundation
OER*	Office of Energy Research
OHER*	Office of Health and Environmental Research (OER)
OMIM	Online Mendelian Inheritance in Man (Johns Hopkins Medical School)
ORNL*	Oak Ridge National Laboratory, Oak Ridge, Tenn.
PACHG†	Program Advisory Committee on the Human Genome (NIH)
PCR	Polymerase chain reaction
PFG	Pulsed-field gel [electrophoresis]
RFLP	Restriction fragment length polymorphism
SBH	Sequencing by hybridization
SCRI	Supercomputer Computations Research Institute (Florida State University)
STM	Scanning tunneling microscopy
STS	Sequence tagged site
UTHSC	University of Texas Health Science Center
VHL	Von Hippel-Lindau disease
YAC	Yeast artificial chromosome

HGMIS MAILING ADDRESS

Betty K. Mansfield
Human Genome Management Information System
Oak Ridge National Laboratory
P.O. Box 2008
Oak Ridge, TN 37831-6050

Comments:

Human Genome Management Information System

Subscription/Document Request (Vol. 2, No. 2)

1. ___ **Human Genome News**
   ___ New Subscriber ___ Change of Name/Affiliation/Address ___ Drop Name from mailing list
2. ___ **Human Genome Information Database**
3. ___ **DOE Human Genome 1989–90 Program Report**
4. ___ **Understanding Our Genetic Inheritance, The U.S. Human Genome Project: The First Five Years, FY 1991—1995 (Joint DOE-NIH 5-Year Plan)**
5. ___ **DOE Human Genome Coordinating Committee Meeting Minutes**
   ___Back Issues ___Add name to mailing list

Please type or print. Enclose business card, if possible.

Name: _____ **Affiliation** _____
(First) (MI) (Last)

Address: _____ _____

_____ **E-Mail Address:** _____

Phone: _____ **FAX No.:** _____