Department of Energy
Washington, DC 20585

August 2, 1991

Mrs. Helen Simon
Chief, Program Planning & Eval. Branch
National Center for Human Genome Research
Bldg. 38A, Room 616
Bethesda MD 20892

Dear Dr. Simon

On September 4-5, 1991, the human genome programs of the Department of Energy (DOE) and the National Institutes of Health (NIH) will hold their third annual retreat. The major purpose of these retreats is to evaluate program status and progress toward meeting the goals outlined in the joint DOE and NIH five year plan. Discussions include obstacles to reaching the goals and possible solutions, but also include the impacts of other activities both in this country and abroad on our planning and progress. It is to this end that we are inviting you to participate in this year's retreat.

While it is necessary to have an agenda or discussion guide to give the group some direction, the actual sequence of topics and issues addressed generally remains flexible. Attached is a tentative discussion guide.

We plan to leave the morning of the 4th open for visits by interested attendees to local area biotechnology companies. Companies with known interest in some aspect of the human genome project are being contacted to determine their interest in hosting attendees. They are also being invited to give a brief (10 min) summary presentation on what they see as problems in the transfer of technology from research labs to the private sector for commercialization. It may be that time constraints will not allow all interested companies to be placed on the agenda. The names of companies expressing an interest in being visited and a contact person will be forwarded to you so you can arrange your visit directly with them.

The retreat will officially convene on Wednesday, September 4 at 2:00p.m. and will continue through the afternoon and evening with a break for dinner on site. This will largely be a time of presentations and discussions. These presentations should emphasize significant progress, but should also address problems and relevant issues that have been encountered and what new steps or strategies are being taken or planned to move the research ahead toward achieving goals of the five year plan in a timely manner.

We will reconvene on Thursday, September 5, at 8:30a.m. until noon when we will break for lunch. The afternoon session is anticipated to continue until 5:30p.m.

I hope you will be able to attend. I look forward to stimulating forward-looking discussions.

Sincerely,

Dr. Benjamin J. Barnhart

Dr. Benjamin J. Barnhart
Program Manager, Human Genome Program
DOE

attachment:

Human Genome Project

DOE/NIH ANNUAL RETREAT
Lafayette Park Hotel
September 4-5, 1991

DISCUSSION GUIDE*

1. Evaluation of Five Year Plan Goals
   - Status-Problems in meeting goals
   - New Initiatives-Facilitation of research goals
   - Sharing Policy-DOE handout

II. New Technology Development: National Labs, Academia, Industry
   - DNA Manipulation
   - DNA Sequencing
   - Data Acquisition

III. International Cooperation-Impacts on U.S. Project
   - HUGO-University affiliation-chromosome workshops; plans
   - Other Genome Programs-Individual countries; EC

*Topics are not exclusive of others and no order is implied.

Dr. Bruce M. Alberts
Chair., Dept. of Biochemistry and Biophysics
University of California, SF
San Francisco CA 94143-0448
[redacted]

Dr. Benjamin J. Barnhart
Manager, Human Genome Program
OHER, DOE
ER-72 F-201GTN
Washington DC 20545-0001
[redacted]

Dr. David Benton
Assistant to the Dir. for Scientific Data
National Center for Human Genome Research
NIH, Bldg. 38A, Room 610
Bethesda MD 20892
[redacted]

Dr. Paul Berg
Willson Professor of Biochemistry
Stanford University School of Medicine
Beckman Center
Stanford CA 94305-5425
[redacted]

Dr. Elbert Branscomb
Biomedical Science Division
Lawrence Livermore National Lab.
P. O. Box 5507, L-452
Livermore, CA 94550
[redacted]

Dr. Charles Cantor
Human Genome Project
Lawrence Berkeley Laboratory
MS 2-300
1 Cyclotron Road
Berkeley CA 94720
[redacted]

Dr. Anthony Carrano
Director, Human Genome Center
Lawrence Livermore National Lab.
P. O. Box 5507, L-452
Livermore, CA 94550
[redacted]

Dr. C. Thomas Caskey
Inst. for Molecular Genetics
Baylor College of Medicine
T-809
One Baylor Plaza
Houston, TX 77030
[redacted]

[redacted]
National Institutes of Health
Bldg. 38A
Bethesda MD 20892
[redacted]5 [redacted]

Francis Collins
University of Michigan
4708 Medical Science II
Ann Arbor MI 48109-0618
[redacted] [redacted]353

Dr. David Cox
Department of Pediatrics
University of California, San Francisco
Division of Genetics
San Francisco CA 94143
[redacted] [redacted]

Dr. Dan Drell
DOE Human Genome Program
OHER, DOE
ER-72 GTN
Washington DC 20545-0001
[redacted] [redacted]

Dr. David Galas
Director
OHER
Dept. of Energy
Washington DC 20545
[redacted] [redacted]

Dr. Raymond F. Gesteland
HHMI-Human Genetics Dept.
University of Utah
6160 Eccles Genetics Bldg.
Salt Lake City, UT 84112
[redacted] [redacted]

Dr. Joseph L. Goldstein
Chairman, Dept. of Molecular Genetics
Univ. of Texas, Southwestern Medical Center
Dallas TX 75235-9046
[redacted] [redacted]

Gerald Goldstein
Physical & Tech. Research Div.
OHER, DOE
ER 74-GTN
Washington DC 20545
[redacted] [redacted]

Bettie Graham
National Center for Human Genome Research
National Institutes of Health
Bldg. 38A, Room 610
Bethesda MD 20892
phone [redacted]

Dr. Mark Guyer
Assistant Director for Program Coordination
National Center for Human Genome Research
NIH, Bldg. 38A, Rm. 610
Bethesda MD 20892
[redacted]

Dr. Diane Hinton
Howard Hughes Medical Institute
6701 Rockledge Dr.
Bethesda MD 20817
[redacted]

Dr. Leroy Hood
Beckman Institute
California Institute of Technology
Pasadena CA 91135
[redacted]

Dr. Elke Jordan
Deputy Dir., Center for Human Genome Rsch.
National Institutes of Health
Bldg. 38A, Rm. 605
Bethesda MD 20892
[redacted]

Dr. Eric Juengst
Dir. Ethical, Legal and Social Implications
National Center for Human Genome Research
NIH, Bdlg. 38A, Room 614
Bethesda MD 20892
[redacted]

Dr. Eric Lander
Whitehead Institute for Biomedical Research
9 Cambridge Center
Cambridge MA 02142
[redacted]

Dr. Leonard Lerman
Biology Dept., 56-743
Massachusetts Institute of Tech.
77 Massachusetts Avenue
Cambridge, MA 02139
[redacted]

Victor A. McKusick
Division of Medical Genetics
Johns Hopkins Hospital
Baltimore MD 21205
phone: [redacted]9

Robert D. Mitchell
Director, Human Genome Center
Los Alamos National Laboratory
CHGS - MS M885
Los Alamos, NM 87545
phone: [redacted]

Dr. Maynard V. Olson
Department of Genetics
Washington Univ. School of Medicine
P.O. Box 8031
4566 Scott Avenue
Saint Louis MO 63110
phone: [redacted]

Dr. Mary Lou Pardue
Health & Environ. Research Advisory Com.
Department of Biology, M.I.T.
Room 16-717
77 Massachusetts Avenue
Cambridge MA 02139
phone [redacted]

Dr. Mark L. Pearson
Director, Molecular Biology
Central Research & Devel. Dept.
E.I. du Pont de Nemours and Co.
Wilmington DE 19880-0328
phone: [redacted]

Dr. Jane Peterson
Chief Research Centers Branch
National Center for Human Genome Research
NIH, Bld.g 38A, Rm. 610
Bethesda MD 20892
phone: [redacted]

Dr. Jasper Rine
Mol. and Cellular Biology
University of California, Berkeley
Berkeley CA 94720
phone [redacted]

Mrs. Helen Simon
Chief, Program Planning & Eval. Branch
National Center for Human Genome Research
NIH, Bldg. 38A, Room 616
Bethesda MD 20892
phone: [redacted]

Diane Smith
Xerox Corporation
4 Cambridge Center
Cambridge MA 01778
phone: [redacted]

---Hoyd Smith
Department of Chemistry
University of Wisconsin
1101 University Avenue
Madison WI 53706
phone: [redacted]

Dr. Sylvia Spengler
Human Genome Center
Lawrence Berkeley Laboratory
1 Cyclotron Road
MS1-213
Berkeley CA 94720
[redacted]15[redacted]

Dr. Robert Strausberg
Assistant to the Dir. for Tech. Development
National Center for Human Genome Research
NIH, Bldg. 38A Room 605
phone: [redacted]

Dr. Robert T. Tjian
Prof. Dept. of Molecular & Cell Biology
University of California
401 Barker Hall
Berkeley CA 94720
[redacted]

Dr. James Watson
Associate Director for Human Genome Rsch.
National Institutes of Health
Bldg. 1, Room 332
Bethesda MD 20892
[redacted]

Dr. Nancy S. Wexler
Pres., Heriditary Disease Foundation
Assoc. Prof., Dept. of Neurology & Psychiatry
Columbia Prebst. Medical Center
New York NY 10032
[redacted]

Dr. Sheldon Wolff
Dir., Lab. of Radiobiology & Environ. Health
University of California
San Francisco Medical Center
San Francisco CA 94143
[redacted]

DOE NIH Retreat Participants

Dr. James Wyngaarden
Foreign Secretary
National Academy of Science
2100 Constitution Avenue, N.W.
Washington D.C. 20418
phone: [redacted]

Michael S. Yesley
Los Alamos National Lab.
MS A187
P.O. Box 1663
Los Alamos NM 87545
[redacted]

Dr. Norton Zinder
John D. Rockefeller University
1230 York Avenue
New York NY 10021-6399
phone: 2[redacted]

Agenda
DOE NIH Retreat
Lafayette Park Hotel, Lafayette CA

Wednesday
September 4, 1991

2:00 p.m.	Retreat Convenes
2:30 p.m	Presentations by Industry Representatives
4:30 p.m.	Presentations (Cantor, Hood)
5:30 p.m.	Break
6:00 p.m.	Meeting Reconvenes
7:30 p.m.	Dinner

Thursday
September 5, 1991

8:30 a.m.	Assessment of 5 Year Goals and International involvement, including role of HUGO
5:00 p.m.	Adjournment

Revised agenda

Agenda
DOE NIH Retreat
Lafayette Park Hotel, Lafayette CA

Wednesday
September 4, 1991

2:00 p.m.	Retreat Convenes
2:30 p m	Presentations by Industry Representatives Affymax Beckman Instruments, Inc Genentech, Inc Genomyx Corp Intelligenetics, Inc. Pharmacia LKB Biotechnology SUN Microsystems Sybase Corp
4:30 p.m.	Presentations (Cantor, Hood)
5:30 p.m.	Break
6:00 p.m.	Data Sharing and Access (D. Galas, E. Branscomb)
7:30 p.m.	Dinner

Thursday
September 5, 1991

8:30 a.m.	Evaluation/Revision of 5 Year Goals (P. Berg, D. Galas)
Afternoon	International Involvement/HUGO (C. Cantor, D. Hinton)
	cDNAs YACs HGM Meetings
5:00 p.m.	Adjournment

DOE NIH RETREAT

SEPTEMBER 4-5, 1991

M. J. KELLY, PH.D.

INTELLIGENETICS

HISTORY

- FOUNDED 1981
- BECAME JOINT VENTURE MAY 1986
  60% AMOCO TECHNOLOGY, 40% INTELLICORP
- BECAME WHOLLY OWNED BY AMOCO TECHNOLOGY 1990
- 59 EMPLOYEES
- BIONET COOPERATIVE AGREEMENT - 1984 - 1989
- GENBANK CONTRACT 1987 - 1992

MISSION

TO PROVIDE SOFTWARE AND HARDWARE TOOLS FOR BIOMEDICAL RESEARCH AND CLINICAL DIAGNOSIS.

AREAS OF EXPERTISE

- WORLDWIDE ON LINE TELECOMMUNICATION SERVICES
- ANALYTIC SOFTWARE FOR DNA AND PROTEIN SEQUENCE ANALYSIS - IG-SUITE, PC/GENE, GENEWORKS
- ACQUISITION TOOLS FOR DNA SEQUENCING
  1. SPEEDREADER AND OTHER IMAGE ANALYSIS TOOLS
  2. GEL, ASSEMGEL AND A NEW SEQUENCE MANAGEMENT PACKAGE
- NETWORKING AND COMMUNICATION PACKAGES ON UNIX BASED WORKSTATIONS

INTELLIGENETICS

MAJOR GOAL

TO ASSIST IN ANY WAY POSSIBLE THE N.I.H. AND D.O.E. TO ACHIEVE THE HUMAN GENOME PROJECT IN LESS THAN TEN YEARS.

ACCOMPLISHMENTS TOWARD GOAL

1. A NEW SET OF NUCLEIC ACID SEQUENCE ACQUISITION TOOLS

   A. ARIAS* - AUTOMATED RADIOAUTOGRAPHY IMAGE ANALYSIS SYSTEM

      - 15 FILM ACQUISITION-AUTOMATIC
      - SEMI-AUTOMATED ANALYSIS
      - HIGHEST RESOLUTION OF ALL IMAGE ANALYSIS SYSTEMS
      - HIGHEST RANGE OF OPTICAL DENSITY
      - AVAILABLE EARLY 1992

   B. ASAP* - AUTOMATED SEQUENCE ASSEMBLY PACKAGE

      - HANDLES MEGABASES AND LARGE NUMBERS OF FILES
      - FASTER MERGES OF FRAGMENTS WITH SPECIAL ALGORITHMS
      - FUTURE RELEASES WILL HAVE HISTORY AND MULTIPLE PROJECT CAPABILITY
      - ATTRACTIVE USER INTERFACE

   C. HIGH SPEED ON LINE TELECOMMUNICATION SYSTEM

      - HIGH SPEED FASTA, FASTDB AND BLAST DATABASE SEARCHES - EMAIL REQUESTS
      - FTP DOWN LOADING OF DAILY UPDATED GENBANK DATA

*CODE NAMES

IG BETAGEN - TROPIX

GENOME SEQUENCING STRATEGY

CHEMILUMINESCENT DNA SEQUENCING ON A COMPUTER CONTROLLED ROTATING NYLON MEMBRANE ELECTRO-PHORETIC APPARATUS (AUTOTRANS - 700) WITH SEMI-AUTOMATED IMAGE ANALYSIS OF AUTORADIOGRAPHS (ARIAS) AND CONSENSUS SEQUENCE MELDING USING A HIGH SPEED MERGING ALGORITHM (ASAP)

GENOME SEQUENCING

THE IG-BETAGEN OPPORTUNITY

HUMAN GENOME PROJECT GOALS:

- IN FIVE YEARS TO SEQUENCE DNA AT $0.50/BASE

IG-BETAGEN TROPIX GOAL

WITH TWO AUTOTRANS 350 @ $10,000 =	$20,000
1 - SPEEDREADER OR ARIAS	$26,000
TOTAL INSTRUMENT COSTS	$46,000

DAILY THROUGHPUT = 11,200 bases/day

COST PER DAY INCLUDING ONE TECHNICIAN = $454

COST PER BASE = $0.04

MAXIMUM COST TO INSURE ACCURACY = $0.12/BASE

Cost Per Base Calculation (IntelliGenetics/Betagen/Tropix)

Hardware Costs

2 - AutoTrans 350 @ $10,000	$20,000
1 - SpeedReader	$26,000
Total	$46,000

Time per Run (hours)

AutoTrans	4.5
Membrane processing	1.5
Film exposure / processing	0.5
Film scanning / editing	1

Personnel

Technician (including fringe)	$35,000

Operating Expenses - Supplies (per run)

Electrophoresis reagents	$3.50
Membrane (6" x 18")	$18.75
Sequencing reagents	$25.00
Chemiluminescence reagents	$25.00
Films	$6.40
Total	$72.25

System Performance

7 reactions / gel
400 bases / reaction of readable, reliable sequence

DAILY THROUGHPUT

4 runs per day (2 runs / machine x 2 machines)
[Morning runs are exposed, scanned and edited the same day. Evening runs are exposed, scanned and edited the following morning.]

4 films x 7 reactions x 400 bases = 11,200 bases / day

COST PER BASE

Personnel (per day)	$140
Reagents and supplies (4 runs / day)	$289
Amortized hardware (5 year amort.) per day	$25
Total cost per day	$454
Total throughput per day (in bases sequenced)	11,200

Cost per base	**$0.04**

SEQUENCING AND MAPPING

CONVENTIONAL RADIOACTIVE METHOD	CONVENTIONAL TIME	IG TIME	IG STRATEGY
1. SAMPLE PREPARATION			1. SAMPLE PREPARATION
2. ELECTROPHORETIC SEPARATION	12-24 HRS	5 HRS	2. ELECTROPHORETIC SEPARATION AUTOTRANS - 700 (BETAGEN)
3. SAMPLE DEPOSIT ON NYLON MEMBRANE	2-3 HRS	---	3. INCLUDED IN STEP 2
4. X-RAY FILM DEVELOPMENT OF IMAGES	12-24 HRS	0.5 HRS	4. CHEMILUMINESCENT TAGS TO DEVELOP X-RAY FILM IMAGES (TROPIX)
5. MANUAL READING OF IMAGE	1-4 HRS	0.4 HRS 0.3 HRS	5. USE SPEEDREADER OR ARIAS
6. REPEAT READING	1 HR	---	6. NOT REQUIRED
7. MERGE FRAGMENTS OR CALCULATE MOL. WT.	1 HR	10 MIN	7. USE ASAP UP TO 2 MEGABASES
TOTAL TIME -	29 - 57 HRS	< 6 HRS	= 5 - 10 FOLD IMPROVEMENT

THE GOAL FOR SEQUENCING IS MET FIVE YEARS EARLY

IG-BETAGEN-TROPIX SEQUENCING APPLICATION	CONVENTIONAL MANUAL SEQUENCING
• 5 HR ELECTROPHORESIS	• 12-24 HRS ELECTROPHORESIS
• USE OF CHEMILUMENSCENCE	• USE OF RADIOACTIVITY
• GOOD BAND SPACING	• POOR BAND SPACING (COMPRESSION)
• BAND DEVELOPMENT IN 10 MINUTES	• BAND DEVELOPMENT IN 7-24 HRS
• OVERALL THROUGHPUT - 11,200 BASES/DAY	• OVERALL THROUGHPUT - 800 BASES/DAY
• COST PER BASE - $0.04-$0.12	• COST PER BASE - $1-2
• 2.8MM BASES/YEAR/TECHNICIAN	• 0.2MM BASES/YEAR/TECHNICIAN
• 1071 TECHNICIAN YEARS = 1 HUMAN GENOME	= 15,000 TECHNICIAN YEARS
• EQUIPMENT COST = $49.3MM	?
• LABOR COST ($45K/TECH) = $48.2MM	• LABOR COST =$675MM
• TOTAL COST = $360MM @ $0.12/BASE	

Networking challenges for the Human Genome Program

Data communication and network connections

Transport and update mechanisms

Local information storage and retrieval

Security and privacy of information

Data communication and network connections

Throughput on the Internet

- T1 (approx 1.5 megabit/second) service
  —Effective transfer time 80 minutes for 100 MB file

- T3 (45 megabit/second) service available now
  —Effective transfer time 5 minutes for 100 MB file

- Ethernet — standard local network for Suns (10 megabit/second); other technologies available (token ring, token bus,FDDI, UltraNet)

Connections via campus/site backbone, or direct connections via regional research networks (or ANS)

Transport and update mechanisms

Manual file transfer (File Transfer Protocol — FTP)

- Support for arbitrarily large/binary files
- Supported on all Internet hosts

Enhanced file transfer

- Scripts/programs to invoke FTP at scheduled times or depending on availability of data

Automatic database update mechanisms using transaction-based protocols

Mail-based file servers/clients

Local information storage and retrieval

Mass storage media

- Magnetic disk technology (SCSI-2, IPI)
- Optical disk media (CD-ROM, WORM, M-O)
- Backup and archival media (Exabyte, DAT)

Information retrieval

- Flat file management
- Relational database management systems (RDBMS)
- Network file systems and local information sharing

Security and privacy of information

Authentication of sender of information

- Checksum schemes (CRC, SNEFRU)
- Formal authentication schemes
  Kerberos (MIT), RSA digital signatures

Data privacy

- UNIX password protection on host
- Encryption (DES, software DES,
  public-key systems)
- Access control lists (provided by RDBMS)
- Privacy-enhanced mail (PEMS) - new Internet standard

End-to-end data security

- Audit trails
- RDBMS transaction logs

IMAGENETICS

sister company in Illinois Anaco Technology Co. also -

- WHOLE CHROMOSOME PAINT PROBES

  GENE MAPPING
  TRANSLOCATION CHARACTERIZATION

- LOCUS SPECIFIC IDENTIFIER PROBES

  GENE MAPPING

DOE - NIH Retreat #3 9/4/91

DG: Org. change. Management at higher level. Dave Smith will be the main responsible person for DOE HGP. Principal point of contact.

Paul Silverman (Beckman)
Interactive program with Human genome centers; adoption of Beckman automated workstations. (Biomek 1000)
Survey: rank order of stated need: libraries, DNA prep, amplifications,electrophoresis,etc
production of high density arrays.

Excellent cooperation with centers
Intellectual property agreements successfully negotiated
Confidentiality agreements: established with an institution, often not necessary (only one in effect at moment)

Benson (Genentech):
Aims of tech transfer: to optimize public benefits, to optimize commercial incentives
Benefits: new drugs, employment opportunities, expanded tax base, expanded exports, royalties for research support.
Ergo: virtue of patents
avg cost of new drug: $230 M
10% is discovery cost, rest is development.
HGP should foster patenting of all new functional DNA information.
Berg:[Issue is one of importance but it wasn't on the agenda!]

Watson: contentious issue. C. Venter wants to patent all cDNAs
Europeans think cDNAs shouldn't be patented. Benson's view is a minority. Depends on a judge.

Tom Brennan (Genomyx)
trying to sequence larger fragments; reduce overall work.
anticipates 3.5 Mb/yr @ $.50/bp
Capillary mass spec approach. Also SBH approach; 10 mer array
Generate high density arrays on borosilicate coated glass. Hydrophilic surface, coat by surface tension. 50 um droplets, 1000 drops/sec. Aim is complete array of 10 mers per hour. Presently, it works with one nozzle.
Status of mass spec: (PB question): background problem (Sulfur)
TB: "give me another two years"

Steve Fodor (Affymax) primarily a drug co.
Light directed synthesis of polymers on solid substrates; photolithography (Science article) Mostly used to date to grow peptides. 256 x 256 array
Light directed oligonucleotide synthesis: chemistry exists. 32 chemical steps would be enough to synthesize all the octanucleotides on a 1.8 cm x 1.8 cm matrix.
Interest in genetic diagnosis.
Concerns: Affymax can contribute to sequencing objectives
funding availability, collaboration, legal issues of ownership, commercial availability,

Kelly (Intelligenetics) wholly owned by Amoco
Ran Bionet. Took over GenBank in 1989. Sequence info now only 2 weeks behind. (See handout)
Processing 100 Kb rapidly, on Sun station, per minutes?
Currnetly working with C. Venter (beta testing) Real data in 2 months
Elbert: can I get to your database? response: yes !

Walko (Pharmacia)
Automated laser fluorescence sequencer. Cosmid sequencing and thin gels
Sequencing: up to about 1000 bp
Thin gels: 400 bp in 218 minutes (400 bp in 3 2/3 hrs)
Caskey: easy instrument to work with, some start up problems with base calling but this has improved with new software. Instrument has very good applicatons for closure strategy, but may not work well for walking.

Colton (SyBase)
(510) 596-3500 proper phone no. Other one is a FAX
Release 5, due next year
ReplicationServer: deals with replicated subsets of data
lots of distant labs creating data, but feeding into a central "home"
Due sometime in 1993.

Charles Cantor talk on DOE technology development. 9/4/91

25 - 33 % of DOE HGP funds spent on tech development.
In the schema, Cloning - Sequencing - Analysis, what are the rate limiting steps?
Principally, it is the "front end" e.g. obtaining/purifying DNA
Needed improvements:
- getting the DNA
- sequencing
- automated assembly/analysis
- cross references between libraries
- efficient pooling strategies
- rapid polymorphism detection
- purified chromosomal DNA

Pooling strategies: most are inefficient and the math behind them is not well developed.

PROBES	TARGETS
1	array of samples
1	array of pools of samples
>1	array of samples
>1	array of pooled samples

A role for DNA triplexes: triplexes are stable at pH 5 - 6, but unstable at pH 8 - 9. Can be used to purify specific target dsDNAs. One experiment (LBL?) $(TC)_{45}$ containing plasmid mixed 1 part in 200 with $(UC)_{19}$ containing plasmid. Using one triplex identification/chromatography step, 99.9% purification of the $(TC)_{45}$ plasmid was obtained.
Lawrence Livermore uses a high speed chromosome sorter, based on FACS, capable of analyzing 20,000 chromosomes/second and sorting 250 - 350 chromosomes/second. This is a 10 x improvement.

Sequencing:
What is needed is optimization of exisiting methods to give a 10 x increase in speed. Also, entirely new methods and new approaches are needed. High throughput methods with direct coupling to databases and analysis are necessary. The needed accuracy varies somewhat with the application.

George Church's multiplexing:
direct transfer from gels to analysis matrix (nitrocellulose); he reports less than 0.6% errors out of 800 bp. He runs 24 reactions per gel and can reprobe each transfer up to 28 times. Using chemiluminescent techniques, he achieves 3 x faster exposures than with conventional autoradiography. It is claimed that 2 x 10^6 raw DNA bases/9 hours can be analyzed.
Oak Ridge, Tenn.:
rather than fluorescence, they use mass spectroscopy (more from Lee Hood to follow) Here, mass spec is used as the detector. stable metal isotopes

(avoiding the use of Sulphur) are used as labels.

Not all tech development has to be "high" tech. Lloyd Smith uses very thin (10 - 100 um) gels (which can be cooled much more efficiently) and very high voltages (9000V over 30 cm length) to achieve excellent separations for sequencing. This is an example of optimization of current methods.

What may the future bring?

- scanning tip microscopy
- single molecule degradation and sequencing
- sequencing by hybridization (SBH)
- mass spectroscopy
- coherent X-ray scattering (this is known now to be an unlikely technology)
- enzymatic groove scanner (based loosely on exploiting what polymerase molecules do naturally)

Problem with microscopy methods is throughput; they are slow. Single molecule scanners would involve tethering the DNA molecule and then successively cleaving the terminal (labeled) base for analysis. Here, there are concerns about the signal to noise ratio.

SBH: tremendous redundancy; this approach is unexpectedly robust. One concern is the problem of end mismatches, but various tricks (polyinosine tails, etc.) may ease this. The real discrimination of SBH, therefore its power, occurs not at the hybridization step but in the washing step. Success or failure may depend on the finicky nature of short oligomer hybridizations

Possible solutions: a) I_4-NNNNNNNN-I_4 (flanked short oligos)

b) NNNN-X_n-NNNN (interrupted short oligos)

Estimates of sequencing rates:
Short (unscientific?) poll. Results derived from respodents only. Raw Kb per person per day (highest and lowest estimates were discarded)

	1993	1996
G. Church	80-240	4800
R. Gesteland	17	-
L. Smith	54	200-300
Drmanac, etc.	300	1000-2000
Jacobson	-	500
"Average"	118	833

For 1996, let's say 800 Kb/person/day. Let's say this is 100 x too optimistic; then it's 8 Kb/person/day
This is about 2 x 10^6 Kb/person/year
If a person "costs" $200,000/year, this is about 10 cents/bp
Conclusion: sequencing is not the rate limiting step; it is, for now, the "front end", i.e. the preparation and purification of the DNA

Informatics:

LLNL contig browser

Oak Ridge: Uberbacher's coding recognition module using neural network approach. Software that approximates the ability to learn. Early results indicate that it works surprisingly well.
This software can detect exons as small as 50 bases.

Near future: Biological Information Signal Processor (BISP) chip from Hood and Hunkapiller. Chip incorporating design features optimized for sequence comparisons. Smith and Waterman approach.

Lee Hood on mass spectroscopy.

principle: volatilize the substrate, measure the mass:charge ratio
Make an eletrospray, feed DNA fragments through the fine spray and put charge on the droplet. Reduce the droplet so the charge is left on the fragment. For protein sequencing, results have been spectacular (at least for small peptides of 25 - 30 residues, in picomolar quantities)

Matrix assisted laser desorption: float fragments in a matrix, shoot laser at the matrix to scatter the fragments or subfragments (get results on larger sized fragments)

Ion Trapping

Time-of Flight Mass Spec: use with laser assisted desorption. Mass of fragment measured by the time of its flight. This can be good for looking at large molecules (pioneer: Don Hunt) At present, the accuracy is not great. This should lead to sequencing rates of 500 Kb/person/year/instrument.

Data Sharing policy

Galas: DOE's guidelines. The issue of data sharing and materials access is a longstanding one and needs resolution. It is acknowledged that the issue of patenting results will impact on this in ways not easy to anticipate now. DOE's guidelines recommend a six month maximum before access to DOE generated results and materials is opened up.

Mark Pearson: conversation with Nat Sternberg, Sternberg says he's being overhwelmed with requests for materials.

Moyzis: still worried about the 6 month period.

Olson: Dollars, per se, cannot solve this problem. It is also a major diversion of personnel. A case-by-case assessment is needed.

Carrano: There are outs in this policy; one could make requestors physically come to the lab and actually do the work. There is a transient problem of freezer space (Chromosome 1 would require about 2000 microtiter trays to distribute)

Olson: no problems with the proposed policy. HGP must be and be perceived to be a source of useful materials to the research community.

Moyzis: The major players in this game are being supplied with what they want, as is. However, the National Labs are not repositories.

Olson: it is the knowledge that is useful, rather than the libraries per se.

Galas: Regarding the sharing of sequencing and mapping data, it is proposed that these guidelines be adopted by the Genome Program as a whole.

Watson: agree in principle. We should use our power as granting agencies to enforce openess and fair access. Other countries will have different rules e.g. the UK on cDNAs.

Caskey: A set of rules is good because it will curtail unnecessary discussion, which is getting repetitive. The momentum of the science will solve problems.

Resolution: Place issue of these guidelines on the agenda for the January Subcommittee meeting (Irvine, California) and publish in the next issue of the Human Genome News with a solicitation for comment.

Retreat 9/5/91
Reconsideration of 5 year goals

David Cox: good 10-15 cM map (average), with big holes. Chr. 21 map is 6 cM average. NIH index marker effort; concern about building on already identified markers. Using CA repeats to place new markers. Index markers very important. PCR based markers.

Mark Guyer: There is a scheduled Oct 6 meeting of index marker people. Things are going well. Chr 4 is in quite good condition; 15 markers are in hand. These markers were mapped using CEPH families. Chr 4 is complete except for two gaps. Chr 11, 21, 19 also in good shape.
lots of activity, hard work. Analagous project in Europe (EUROGEM).
Goal: finish maps by Sept. 1993

Tom Caskey: Simple repeats will be the markers of greatest utility. However, triplets and higher order repeats will be even more useful. New sequencing technologies will make a difference also. Caskey sees value in a heterogeneity of approaches, but no real problem achieving the goal of a human map with markers every 2 - 5 cM in 5 years.

David Cox: Chr 21 has lots of markers but only 5 have heterozygosities greater than .7

Lee Hood: automation will allow use of sequence directly for defining markers. Over 40Kb of sequence, he sees diallelic polymorphism every 500 bases or so. 96 well microtiter plate assays are practical. Sequence based markers are the future. PCR + dialleles, rather than simple repeats, are the future.

Caskey: With organization, goal 1 is a cinch.

Leonard Lerman: how will progress be measured? A histogram?

Benson: "report card" as histogram; a picture by January?

Watson: have mapping subcommittee meet, prepare report for January meeting
Hood: we can do the map, but the technology is not widely dispersed; genome office should push the technology. At an earlier Salt Lake City meeting, the consensus was that the mixture of technology was not optimal; now, PCR is taking over.

Guyer: the move is away from the use of, or reliance on, CA repeats.

Maynard Olson: CA repeats are still in use

Caskey: avoid using the term "PCR"; prefer amplification based.

Cox: This could be a role for the mapping working group.

Charles Cantor: should we add index markers to goals?

Elke Jordan: HGP doesn't have its full $200M; it would be understandable if the HGP didn't meet every goal.

Cox: 2cM vs. 5cM?

PHYSICAL MAP GOAL:

Olson: This was always an ambitious goal. Either we should restrict the goal or dilute the definition of it. It is better to do a few maps well, rather than lots of maps poorly.

Cantor: I agree. For serving the community, incomplete maps are very useful.

Hood: a major purpose of this goal is to drive the technology. The problem of closure is a big one. I agree with Olson.

Elbert Branscomb: What is the definition of closure?

Tony Carrano: The utility of maps is important.

Bob Moyzis: This goal is achievable, except that 50% of the genome is not being worked on. Should we work on a lower resolution map of the other parts of the genome?

Galas: The definitions are very ambiguous. What is "large parts?" What is completeness? How are the maps to be used? These issues are technology dependent.

Hood: Mapping is expensive. The options are low grade efforts on all, or high grade efforts on some.

Olson: Several concurrent technologies, all of about 1 - 2 Mb sensitivity, exist; thus "agreement" cold be on 2 Mb resolution.

Paul Berg: Are Sequence Tagged Sites (STS) markers realistic?

Olson: The goal is OK, we should stay the course.

Moyzis: There is no valid scientific reason not to achieve this goal; economics is another matter (5 year goal should become 10 year goal?)

Cox: What is the goal, the map or the clones from the map?

Carrano: What is continuity? A map with small gaps is still very useful.

Berg/Hood: The second sentence "Generate overlapping..." is redundant.

Watson: There should be no gaps.

Carrano: Gaps are not so important.

Watson: Carrano (at the last joint meeting, 6/25/91) declared victory.
(Carrano: no, we haven't)

Hood: There are unclonable regions, which will be a problem.

Berg: What about assembly of an STS map?
Hood: we should aggressively collect STS in a database.

Watson: Let's look at the genetic map in Jan, 1992 and the physical map in Dec. 1992. We can compare Carrano vs. Olson in Dec. 1992
We must not forget the Japanese presence. STA appears to be taking the lead role in the Japanese Genome effort; 40 Japanese were in London They want to do Chr 3, 11, 21. Japanese want to do what we're doing; like automobiles, they could do it better.

Caskey: This is the most boring part of the whole genome project. We need to look at a variety of approaches.

Galas: Using a variety of approaches towards defining physical markers is worthwhile. Discussion indicates there is ambiguity about this goal. How many serious efforts are now underway towards physical mapping of different chromosomes?

SEQUENCING (Goal 3) 10 million base pairs

Watson: With regard to Drosophila: DOE pays for tech development, NIH pays for biology "DOE relatively richer than NIH"

Hood: Having a number of groups doing this is critical. Lots of commercial companies are doing small scale sequencing, none doing large scale. Such
groups need to be recruited.

Gesteland: I'm planning a large scale sequencing course.

Hood: Using model organisms would be better for the 10,000,000 contiguous bp goal (human has lots of repetitive sequence) Large scale sequencing is better done by dedicated enterprises (less likely to be academic)

MODEL ORGANISMS:

Elke Jordan: model organisms are a sequencing issue, except for the mouse.

Watson: Fred Blattner tried a major effort on E. coli, it was an unmitigated disaster. He wanted 1 million final bp, he got 1/10

of that. He got some critical informatics help from a "sensible" Israeli. He was given an ultimatum: show up at CSH E. coli meeeting (6 wkks) with a manuscript on his E. coli work. George Church is only interested in technology, and is thought to have the E. coli sequence but it's on a stack of "unreadable" films. I've always thought of E. coli as an American organism. More people are needed in the sequencing game. The model organisms effort will succeed if DOE gets involved. Maybe an aggregate 10,000,000 bp is a more realistic aim than 20,000,000.

Hood: Involve companies. Hoffman La Roche's involvement in PCR may lead to a patent issue for genetic testing as it may pertain to sequencing. The future is diagnostics; this will be problemmatic if its PCR based.

Lloyd Smith: Applied Biosystems Inc. can't save the world. When giving out a contract, be sure that contracted company has done some preliminary work.

Hood: We want to avoid flaky companies.

Galas: Should we be doing many different model organisms? Should we be focusing on one or two model organisms or many?

Mary Lou Pardue: The reason to do model organisms is the biology. Sequencing alone is not a wise expenditure of effort.

Mark Pearson: We should not restrict which model organisms to work on; let natural selection decide.

Galas: Genome sequencing for technology development can be done on any organism.

Hood: The committment to technology development should be strong factor.

Berg: My summary is "stay the course"

Elke Jordan: There is no current attempt to sequence the mouse genome. The goal is to do the genetic map first.

INFORMATICS:

David Benson(NIH): The first goal is done.

Branscomb: The second goal is in the works; also, yesterday, Intelligenetics said it was all done (part 3). Part 3 is an open-ended goal.

ELSI:

Nancy Wexler: The nice thing about the ELSI goal is that no one will ask for a histogram! Entreprenerial drive is for genetic

testing. 2 sets of goals
1) research agenda (both NIH and DOE)
DOE emphasizing privacy and confidentiality
NIH: CF
Other areas where more "proactive" approach is needed:
1) delivery of genetic services
2) IOM/NAS project on genetic testing/services (Conflict Of Interest problems apparently solved)
3) privacy (Yesley privacy planning meeting) pre workshop next week
How specific does one want to be about genetic info vs. medical info. HGP as catalyst for related issues.
4) insurance task force
5) discrimination, esp. in employment (EEOC eviscerating Americans with Disabilities Act). Berg/Wolfe letter to EEOC. Response: "we never intended that, it isn't a problem, it isn't relevant"
Discriminations can target three types of disabled: the actually disabled; those who once were disabled, but were treated; those who appear to be disabled.
Can "genetically disabled" be defined??
Elizabeth Thornton wrote the response to the Berg/Wolfe letter. People with late onset genetic diseases but who are not actually symptomatic are protected under ADA. (Wexler more positive about this than others)[should be clarified]
Berg: What is the ELSI working group doing to deal with unrealistic expectations from general public? Didn't Watson say, at Senate hearing, that main object of the HGP is to identify human disease genes?

Wexler: Finding the gene is just the first step; much follows. Education, interpretation of results.

Watson: The general concern for ethics is on the rise. Congress wants it done.

Elke Jordan: NCHGR is going to spend 5% of its budget on ELSI next year.

Caskey: genetic tests are very precise; not like other medical test areas; CF may be only 92% accurate, but other tests are much less so. It is necessary to keep a balanced perspective.

Norton Zinder: ELSI activities are to study, consider, recommend. They are NOT to write law.

Cox: This goal (ELSI) has already succeeded.

Wexler: A remaining problem is how to deal with large kindreds.

TRAINING:

Bettie Graham: NCHGR has made 4 predoc awards for a total of about 30 students. Awards have been made to 6 minority pre doc

students. 20 postdocs are in training (mostly in molec. biol.) NCHGR has not been so successful for other disciplines. SIRCA: for mid career people ($50,000 per year salary, $20,000 per year for support.) Announcement due end of September. Availability of courses. 6 courses (1 in ethics) exist; it is a regular program now. Basically, the training goal is being approached, but slowly.

Mark Pearson: The HGP program is not so successful with interdisciplinary candidates (engineers, physicists, computer people, etc.)

Galas: DOE has 10. The centers are the places for these, primarily.

Hood/Watson: broaden fellowship sites to med schools?

TECHNOLOGY DEVELOPMENT:

patenting?

Bob Strausberg: What should HGP be doing that it's not?
Watson: The ideas/proposals being turned down are bad ones; they won't work. Study sections are inherently conservative.

Strausberg: R21 mechanism is for pilot projects.

Lerman: Are tech development proposals being recieved from academics?

Elke: That's most of them.

Jane Peterson: P20 mechanism. This is an interdisciplinary group mechanism. Tech development can play a role here. Many proposals poorly written.

Elbert Branscomb: Regarding poor grantsmanship. NCHGR should help.

Diane Smith: on date of approval, FOIA can be used to gain access to grant. PI can mark pages as proprietary, but if he/she doesn't, it can be released.

TECHNOLOGY TRANSFER:

Berg; This is a platitude, easily fulfillable.

Galas: It seems to be the sense of the group that the goals should stand as they are.

Zinder/Berg: We're only one year into the clock, so it's too early to be changing the goals.
Galas: Let's revisit them in a year.

Hinton: HUGO. Council meeting in London. Wyngaarden, Director of HUGO International, has resigned. HUGO searching for a replacement, leading candidate has been identified. US office: housed at HHMI in Bethesda. HHMI support stops this month. $700,000 in bank right now (not very much) Negotiations with Johns Hopkins University are "active". European office is in London. Japan: Science and Technology Agency is taking the lead (this should increase efficiency). Moscow office. Interest from NATO; October meeting of NATO with Diane and Bronwyn Loder. Significant struggle: the purpose of the organization.

Shelly Wolfe: coordination was supposed to be the strong point of HUGO; how is it dealing with the Japanese?

Hinton: Japanese HUGO representative is often not the significant person; the relationship of HUGO with Japan is different from the relationship of HUGO with anyone else (so far; this may be changing). HUGO has been more successful with other countries. HUGO started as a small organization, of scientists, to coordinate and provide scientific advice to governments (there was concern that governments would be making scientific decisions they were felt not to be competent to make). HUGO is now looking at chromosome workshops and an international ELSI program (Wexler) and a study/commission on international property.
Cantor: HUGO started out as a fairly elitist organization; it is now changing character. For a long time, HUGO was dealing with the wrong people in the wrong way vis-a-vis the Japanese. When HUGO started, there was a preexisting Human Gene Mapping organization. This organization outgrew itself. The latest meeting in London cost at least 2 M pounds (possibly more).

Hood: In some ways, HUGO has been a failure.

Wolfe: HUGO's role is that of a middleman without a budget.

Hood: HUGO has possible roles in training, ELSI, Third World.

Cox: The death knell of HGM meetings was the introduction of real time computer data entry. Fix meetings by removing computers.

Cantor: 3000-4000 cDNAs have been done in Europe, but none are in any database. This is MRC policy, but the whole of the EC has bought off on it. cDNAs are free to academics, but there is a subscription fee to companies. This is specific to the cDNA project at MRC (Brenner) and associated labs.
HUGO proposes: collect information on cDNAs and the people doing them; organize some sort of database searching protocol; exchange charcterized cDNAs (to avoid duplication). There was no representation from Japan at this meeting. The European program has not made loud noises about high resolution mapping.

GDB

Watson: Regarding participation of other countries: a letter was sent from Watson and Galas; the response was "tell us more." There will be a Dec. 16, 1991 meeting in Paris. As to funding GDB, a possible breakdown is: 40%:40%:20% US: Europe: Japan. There are to be 2 advisory groups, one governmental, the other scientific.
(Bodmer: make HUGO and GDB synonymous)
Is there a role for HUGO in GDB???

Watson: HHMI is upset with HUGO because HUGO wasn't going to do any ELSI.
Europeans only hold meetings in resorts!

Elke: like Hilton Head!

Watson: the Executive Committee (of HUGO) is as loved as the Communist Party.

Cantor: on the subject of YACs.
General policies for handlling YAC libraries are different in US vs. Europe. In the US, freezer space permitting, YAC libraries in their entirety are available (with very few strngs attached). In Europe, access to YAC libraries from UK Resource Center, is much more difficult. A central data repository (CEPH model) is the rule. There is a subscription system; prior to access, a promise is required to return any data to the centralized database/source.
Any hits by public domain probes, tested on different libraries, should be public knowledge. How do you know? Exactly the question! Perhaps GDB should serve as database.

Carrano: CEPH library is available in US, not in Europe.

USSR:

Watson: We should do something to help USSR. a formal program for financial assistance to USSR.

Galas: Getting some young Russian scientists over here would be beneficial both for them and for us.

Watson: About 20 would be a good start.
Elke Jordan: Some programs/mechanisms for this already exist.

Watson: Some more formal-looking program should be set up. Something more visible.

Watson: some way of honoring Jean Dausset for CEPH; something here in the US

Wexler: Involve the rest of CEPH as well.

Graham: Dausset Fellowship program? for Russians??

Cantor: regarding the USSR, foreign currency has enourmous impact.

Galas: LANL/Vanderbilt collaboration with Leningrad

Moyzis: there are some outstanding scientists in USSR. Those people are hungry.

[Helen Simon's Notes]

Retreat NIH-DOE 9/4/91

Galas - reorganized DOE -
Hlth Effects + Life Sciences - David Smith
Director now head of genome Program

tech development

Industry Reps:

① Silverman - Beckman - has automated work station; interact w most of NIH Ctrs. Protocols developed w 3 Ctrs; working w Botstein. Good cooperation; intellectual prop Rts worked out w each indiv institution.

② Benson - genentech -

Problems

① not optimize public benefits of human genome prog
② not create commercial ventures

Public Benefits

① new pharmaceuticals
② employment opportunities
③ expanded tax base
④ expanded Export value
⑤ Royalties to support research

New Drug Costs

① about $ 230 m each — 10% Discovery
— 90% Development

Patent laws -

① must be new, useful, non obvious
② composition of matter patent
③ method of making patent
④ method of using patent

Human Genome Program must consider its impact on stimulating development capital

Publication
① raises patent barriers
② uses become more serious.
③ reduces development incentives

HGP has promoted patent protection.

Berg: this was not issue to be discussed. Some confusion - Sylvia may have renamed this. She told companies agenda was issues related to technology transfer.

Watson: issue of patenting came up before Domenici hearing. Craig Venter planning to patent cDNAs. Angen is opposed to this. Was discussed at length at HGM II in London - Europeans opposed to patenting DNA sequences.

Galas - need to organize a mtg soon to discuss whole issue of patenting

③ Brennan – Genomyx – part of Genentech. Dev of seq. method using mass spect. Productivity = 3.5 megabases / yr. Machine = $65-70,000.

④ Fodor – Affymax – primarily a pharmaceutical company; much work on peptides – light-directed synthesis. Oligonucleotide arrays are used for sequencing. Long term goal = genetic diagnosis. Open?: how best to participate in primary sequencing.

Relationships w NIH/DOE: need to understand funding availability; legal? of ownership; commercial opportunities for collaboration.

⑤ Kelly – Intelligenetics – hand out. Major goal – assist NIH/DOE to achieve HGP in less than 10 years. New seq. acquisition tool – available in 1992.

Dr. Watson asked most of the speakers whether they were presenting theory or something in practice. If theory – when in practice?

⑥ Waiko – Pharmacia – ① cosmid sequencing – sequence large segments of DNA; custom primers + vector primers ② Unigels – ↑ throughput (will be available later this year). 4 sites: NJ; Milwaukee; Alameda, CA; base in Sweden.

⑦ Colton – Sybase – 1992 will bring out new Sybase – #5. Will have a much broader selection of software tools.

Problem – cannot store vectors – Sybase is a relational database and vectors are not relational. Studying intensively – probably 3–5 years.

Scientific market is a small one for them.
3070 = Wall Street, banking.

Sybase is small (1000 people), growing. Competing with ORACLE – lge company.

CANTOR: Review of techdev funded by DOE –
Between 1/4 – 1/3 DOE funds in techdev.
1/3 of this = instrumentation
DOE labs use robotic manipulation of their libraries
Not all advances in tech development are expensive and high tech.
DOE has national gene library at Los Alamos + Lawrence Livermore = sole source for many of these materials.
Sequencing – DOE's goals – to optimize existing methods to gain a factor of 10 in efficiency; develop new methods to gain factor of 100 in efficiency

HugE gels = human ultra-thin gel electrophoresis (Lloyd Smith) = new technology of thin gel for faster throughput.

Hood - mass spect holds great promise for seq. technology - ① as a gel reader and ② to substitute mass spect as a gel reader or ③ to take lge fragment of DNA → mass spect and analyze = 3 ways mass spect can intersect with sequencing.

Electrospread: feed fragments, multiple (+) chgs on multiple fragments.

Also mentioned some new technologies that have been developed over the last 10 years for analyzing fragments.

Data Sharing + Access - DOE policy - hand out - Galas presentation on proposed guidelines. Presented as a candidate for genome program in general to adopt.

Pearson? - must address need for flexibility when a lab is overwhelmed by demands more than they can address.

Galas - DOE labs face same problems → flexibility.

in responding + need for additional resources to respond.

Mayzis: discussed these same issues 2 yrs ago. STSs developed by Olson address this problem of ~~at~~ wide spread availability. 6 month is a problem. 5 years from now STSs will solve this. This is a short term problem.

Olson: agencies need to understand that ~~&~~ cannot solve this problem. It involves a major diversion of key personnel who are needed to make progress in genome research. This resource is expensive and we need to be slow in rushing to fill all orders.

Cacano: this lab has gotten around the problem by asking requestors to come to the lab. Also an acute problem with freezer space.

Galas: let Centers make the judgment on usefulness/value of the request. This will work in short term.

Olson: HGP must be a source of useful research materials in a reasonable way.

Mazzis: labs have done a good job of distributing material to the major labs. They are not repositories to large #s of labs.

Olson: technology is much more useful than actual material.

Watson: agree with principles - need to encourage openness. Can restrict funds to those who do not share material.

Galas - goal is not to address individual research cases but to establish guidelines to be helpful

Berg - people who don't give, don't get - they get excluded very quickly.

Watson: data sharing specifically planned for Chromosome 21.

Caskey: these guidelines would be useful in stopping a lot of useless discussion on sharing. Each center can list available sequences in its publications. This will accelerate collaborations.

Cannot vote on it here. Caskey's ctr will pen our guidelines in absence of an NIH policy.

Watson – at jt mtg in January mtg.

Pearson – distribute to community first?

Jordan – we did discuss at ctr directors mtg.

Berg – publish in newsletter + ask for comments

Branscomb – data will go into GDB. Contigs are being transmitted to GDB. Can have users access data bases directly w some restrictions.

Watson ? on how published – when data → GDB the information is anonymous (Schlessinger concurred). If send to a journal, it takes 6–9 months. ? electronic journal NIH/DOE to identify sources of maps + get info out quickly. Perhaps need to have a meeting of those creating data to determine how best to share it.

Caskey: to get this data in journals – mapping data will have to be packaged w function to get it published. Mapping data alone will not be of interest to journals.

Cantor: Could a 1 page summary be prepared to summarize & get a literature citation.

Watson: this trivializes a year's worth of work — not adequate to show your Mother 1 page for all that work.

David Cox: needs more than a single page for important data. Where is vehicle for multiple labs to publish.

Watson: need to seriously address this issue. We need to start a journal. Just need to decide on format & basic approach.

Woolf: if ask for significant citations, will this kind of non-refereed journal have any value.

Gajer: could the criterion be that the data had to have been presented at a single chrome workshop

Cox: cuts people out who can't come & slows it. Validation is very difficult - look for inconsistencies.

Tjian — why can't it be reviewed?

Cox – can't look at the raw data.

Sept 5 Update of 5 year goals 8:30 – 2:45

Genetic Map
① Where are we at end of 1st year?
David Cox: Generally a good map at 10-15 centimorgans but some holes: 21 has best map ($\bar{x}$ = 6 cM); 8 is a mess. Effort this yr is on index markers – will help us identify the gaps. They are not building on markers already there, but starting from the beginning to get a whole new map. They will be critical in the next 2 years.
Key?: those daring markers – some only PCR based, others went to just get them. They are useful for rapid scanning of chromosomes.

Mark: Advice – that would be useful. Planning mtg on 10/6 of markers researchers. Some grants began last April, some a year ago. Will update at Jan mtg after Oct mtg. Preliminary rpts look like going well – C4 = 15 markers well spread out (last Dec = 10 markers), # candidates has doubled. There are only 2 gaps greater than 20 cm. C19 + C17 also in good shape. Other end of spectrum = C10. Index mappers have taken this seriously – going well.

Analogous groups in Europe = Eurogen (23 labs funded – now organized at chromosome basis). Some are being assigned, some will be chosen by researchers. Will be added to US

maps or CEPH maps. C21 + CX will soon have an index responsible for them. -- to finish them off. There are no chromosomes not being covered by index mapping. $\bar{x}$ spacing = 10-15 cm. Goal is finished in Sept 93. Caucasian goal is higher resolution of existing maps.

Caskey - ① believe simple seq repeats will be the technique for mapping.

② w some additional improvements in technology - easier to forward for than retrofit. Not worth a lot of effort to retrofit

③ if go a STR information, or very least get STS + 50% of time get polymorphic info

④ have been able to multiplex with fidelity not possible w CA repeats. (Guyer - ? if high mutation in CA repeats - Caskey - yes, family studies show high rates of mutations in CA repeats)

Goal is easily achieved. Heterogeneity in the community will not serve us well.

Cox: need to be clear with the community with guidelines on best approach.

① PCR based: should be the approach used.

② heterozygosity - more Retro3 - more people useful for. But confidence level is much higher if have more probes - get spacing rather than perfect markers.

Linkage markers will be much more useful to the community if are PCR based than if focus on heterozygosity.

Hood: real future of genetic mapping because of automation:

① will search for polymorphisms by DNA seq analysis ~~Have looked at 10,000~~ Can automate the analysis completely

② C14 studies 500bp → 1.5bp - found mult alleles.

Because of diff in evolution - can create highly informative markers. Can focus in and make the markers highly useful.

Caskey: agree that cannot ever get density wanted. Repeat seq. have more information. It is easier to get info prospectively than retrofit.

Lerman: in evaluating progress - could be clever to see a histogram of markers - compare from yr to yr by chromosomes.

Benton: creating annual score card. Histogram is feasible. Very valuable suggestion. [Berg - when?] Data base design is done, collection will take longer - by Jan mtg.

Watson - need to have mtg of Mapping subcomm to review data before Jan mtg + after Oct mtg.

Zerg - info mtg before Jan -

Watson - don't want this to be a staff report but needs to come from committee.

Hood - agree - worry that techn has not been adapted by many mappers. Imp role of genome office - to push techn and not stay buried in the past. May not reach goal w/out genome office pushing P on mappers to use technology

Cox - mistake to micro manage too much.

Guyer - purpose was to get markers to community as soon as possible - now believe they need to be PCR based.

Olson - history of this very complicated. My impression - most index markers using CA repeats

Guyer - original applics were - many moving on from there.

Olson: Map. subcomm was less enthused about CA repeats than Study Sections that reviewed them. Community is getting mixed messages.

Caskey: could make wiser decisions now than 2 yrs ago. Some systems have clear advantages and should be encouraged.

Olson: if have a mtg, need to be up to our elbows in data.

Berg - hear 2 things
① need for mtg - include discussion of technology
② develop a score card that can develop from year to year.

Alberts - ? of development of technology - will this be developed by companies & our role help.

~~Hoot~~

Cox: from map working gp - talk to people in the trenches of what is working well. Can help staff in assembling data & in making recommendations.

Guyer: past June had grantee workshop - ? do next June for linkage map grantees?

Berg – yes

Cantor – add index numbers as goal

Watson – do not revise goals now.

Berg – goal is longer, not always how we get there.
? Watson – does Congress ask.

Hood – need to discuss level of detail.

Ej: 1 overarching factor – based on budget proj not being met. May be operating at 2/3 of what we planned.

Cox: big diff between ± 5 cm. Need to focus on where errors are – can convey to mapping community

Physical Map.

Olson: most individuals in "key parts of human genome" this is an ambitious goal. State of art moving rapidly during time supposed to be achieving. this is area where budget level being lower capacity but even moving so fast, it & may be OK. Fund decision: maintain ambitious definition + be somewhat specific on few maps. or delete goal. Vote for #1.

Cantor - agree this goal will not be met. But disagree w Olson - incomplete maps are still useful valuable resource. Would endorse progress on all sides.

Hood: agree w Olson - will drive technology by doing things well - come to closure on a few done well rather than sloppy on 20.

Branscomb - need to have a discussion of closure Perhaps not as specific as original goal.

Cavano - orig goal was not clear → confusion in the community. We want good maps - better to have incomplete maps than very best possible.

Mazzi: goal as stated is achievable, but we want ① & ② only 50% of genome is spoken for, not being stu

Is there one need for getting a lower resolution map in the next 5 yrs.

Galas: definitions here are very ambiguous. Need to discuss and/or redefine. Need not choose good or less good map. YAC map or cosmid maps are very different. Need to look at how maps are going to be used. Progress is much quicker, more technology-dependent than genetic map.

Hood: mapping is expensive - need to decide how to focus on priorities - cannot do it all

Galas - when to declare victory - sci issues

Ef - very persistent issue

Berg - need to know what costs will be. Need to discuss mix of 70 mapped carefully vs incompletely.

Olson - goal #2 of 2m base pairs - was judged to be technically feasible and long term assembly of physical maps can be done by several methods. The variety of technologies all have this order of resolution and high resolution maps must be consistent with this.

Watson - my impression is that Europ do not support STSs.

Berg - our own teams are not committed to it

Olson - there is resistance to STSs in Kark + Lele but Centers are generating STSs. Committee needs to set some policy. The case for using STS is overwhelming + need not redo the goal but stay the course.

Mayzis - no sci reason to change goal; economic reasons may mean 10 yrs instead of 5. Yes it is difficult, but if we don't finish the job who will?

Cox: need to not back away from this goal - tough but completely achievable.

Watson / Canano discussion of gaps + age of gaps.
Watson: need to deliver what we promised

Hood: need to remember power of STSs:

Berg - then 2nd goal is redundant.

Watson - will not know for 18 months when reports come in from Centers like Olson.

Caskey/Watson discussion of gaps

Berg – Score card here as well # STSs, where linked to

EJ: that included in what David doing.

Hood: how doing – a zillion STSs that have not been coordinated.

EJ: how assemble

Hood – you set up data base & then go to the community to provide. What you have now is only a fraction.

Olson – measure of progress is critical. Histograms may be very useful.

Watson – year from Jan – lge update on physical maps. genetic map this Jan.
Jan 92 – present what have, contigs but not STSs
Jan 93 – include progress + STSs.

Berg – then miss sense of progress if wait a year to report baseline. Report what have now.

Watson: Japanese are moving in to area. C3, 11, 21
Will see what another nation will do. Must be aware of them – competitor

Watson - Tony is still trapped in cosmids

Cavano - not fair - 90% of community trapped in cosmids.

Watson - Tue look at car analogy - still does not mean it is not shit!

Cavano - We will be held accountable. Congress + the public care less about #STS than #genes.

Casey - this is boring research + need to find ways of ↑ productivity.

Olson: endorsed importance of going after large # of STSs.

Galas - the discussion today demonstrates that this goal is indeed ambitious. Less risk than when we declare victory is where do we go from here. Need to see where efforts really are

Watson - apples came in from all we thought. Could do the job when began this program. Apples now are not very strong. Need to focus on what is going on. A year from now we will know which of diff approaches works best.

EJ: then need to focus on chromosomes we have started, rather than trying to do all.

Berg - then Watson sugg. wait until 1992 to see which approach is best?

Watson - yes - take that year

Alberts - ask 10 of the leaders to write a 5 page document on their position. this is very hard to follow - some real sci issues here.

Watson - not enough info now.

Olson - agree that 1 yr is ambitious goal - something to show the outside world then.

Watson - by this Jan we will summarize data we have.

Pearson - Ctr Dirs have thought about this. In an add'l yr we will be looking to them for progress. We need to dev criteria very soon so are talking apples + apples.

J Peterson - Ctr Dirs mtg in July - would be interest among them of criteria being developed.

Cox: each ctr knows now its forman for releasing data- need to know if that is nor OK.

DNA Sequencing

Hood: if Cantor estimates are accurate, we are in good shape.

Cantor: those came from mouses, but very positive.

Berg: why was cost specified?

Watson: 50¢/base pair → $3 billion cost. And did not want to go ahead with major seq until the cost had been reduced. Lee is behind schedule (Hood disagreed). Now yet at 100,000 bpairs.

Hood: not the issue. With the multitude of approaches, 5 yrs + 100,000 base pairs is achievable.

Watson: WORM group there are producing data. Will be lucky to 4-5M base pairs over 5 years. My belief is most will fail and we will get 1 winner, not the genome. WORM will succeed. ? re Drosophila - may be ready to seriously seq a yr from now. Proposal is NIH pay for seq + DOE

pay for seeders. That is garbage. DOE budget has gone ↑, NIH has not. Hope they will pay for ½ of it.

Hood: unless you have 4-5 gps seriously doing lge scale seq we will not make the goal. We are not behind schedule - can match more people. There will be exponential progress each yr so need to get 4-6 sequencing groups. No groups are really set ↑ to do lge scale seq in industry - only small scale or less.

Watson - we would welcome an application - have encouraged ABI.

Berg - visited ABI yesterday, they were not enthused. They are the only ones able to do large scale. They would need to collaborate.

Watson: need to wait another yr. ABI might participate then.

Zinder: original goal - no major seq until seeders occurred: NIH was a goal in seeders.

Watson: Think we want some human DNA so we know where the problems are. Jim Watson is pressing for sequencing faster.

Ray Gesteland - next summer, 3 wk course on genome tech - hands on seq + informatics. to bring in 16 students (graduate, post docs). Uncertain - how deeply to go into unusual methods.

Bob Strausberg - imp to bring new people into field quickly. Ray ~~White~~ Gesteland will do as part of Ctr. Developing an RFA ~~for~~ courses.

Alberts - 10 m bps is practical. How imp then it be human?

Hood - yes, human is diff; enormous technical challenge w Repetitive DNA - diff from model organisms If we modify goals - might think about models instead of human.

Berg - estimate of 50 ¢ includes front end costs

Hood - as this evolves, those doing phys map will collab w lge scale sequencers. - See over time a joining of Ctrs or participation of companies. There is a large Role that someone will fill. Probably not the academics, but industry.

Watson - need to keep academics involved in the process.

Model Organisms

Ef - most are a seq issue except for the mouse

Watson:

E coli - Blattner has a seq. grant to do it - an unqualified disaster. Gels did not WORK. Sent visit in July. He claimed to do 1M, completed 100,000 this year at best. Hated Elchange. Fred is a tinkerer; not adequate space - space has been promised. Informatics person may bring some order. Ultimatum - must be an CSH. Mtg. on E coli with a manuscript or we will prepare an RFA on E. coli. Fred will likely fail. Church is only interested in tech dev; limits unreadable.

Seq - Elke says Japanese are doing it. To limited funds, why do an RFA on E coli.

Watson: We will not be beaten by Japanese - this is a cottage industry. E. coli is an American organism. We need to get more people into seq - 1 way to get them in is their model organisms to test, then humans.

Yeast: Europ are bribing people to do it by paying more than it costs. 2 approaches being tried at Stanford now by Botstein.

Watson - modeling could succeed if DOE would agree to fund them. This goal is achievable but not by NIH alone.

Hood - to achieve this we would need 10 groups working on it. Even if had $ do not have people. Perhaps goal should be 10 m instead of 20m.

Watson - we would like you (Hood) to 2m bp in 2 yrs. We did an RFA or DROS to respond to Lawrence Berkeley request. Who will pay for production? Discussions w NRC on this.

Hood - we can't interpret if get some companies interested in lge scale sequencing. Industry has many hammers - we need to make it easier for them to enter this field.

Watson: if we encouraged now we could get a lot of worthless applications

Ef: we could encourage a company we have faith in

Berg: could do a cadmack w ABI if they were interested.

Hood: maybe steering group could consider this.

Bob Strausberg - seq gp mtg in Hilton Head 9/26.

Pearson: could invite people to Hilton Head to consider federal investment in seq w tight $: need business people to advise on the future.

Watson: future is in diagnosis.

Hood: serious patent issue involved. Hoffman La Roche has PCR patent + they are interested in diagnosis.

Pearson: Kodak is suing + could be tied ↑ for a while

Hood: Cetus has backed off from extreme position but still tight.

L. Smith: premature to expect ABI to do seq for us. they could run a test before a lot of funds are invested.

Hood: company would need to set aside resources + personnel to this for a time. Need to have an incentive of profit for them. Need to have a demonstration of thruput, etc. to need ack Sc[illegible] proposals.

Galas: may face probl of resource allocation: should we be seq a large # of organisms or focusing are

resources in tight times. What is rationale for a lot of modeling?

Watson: at least 10% of effort = models. Need to do models for gene identification.

Galas: not my?; should we focus on 1 model?

Watson: CSH mtg on Dros.

Pardue: use models in productive way - not useful to just seq Dros genome. Work doing w/ Dros is superfluous. Spending $ on Dros not best use of HGP $.

Pearson - these are 5 yr goals in a program that will take longer than 5 yrs. Many will take more than 5 yrs. Let natural selection take place. Don't do serious weeding at this time. Do not agree w Galas that we should restrict model organisms at this time.

Galas - genomic seq for tech dev can be done on any organism. What are real arguments for multiple models

Hood: support Pearson strongly. need to support those

interested in seq - let evolution take its course. Encourage as many people in any system they use to get sequencing. There are so few people committed to any seq we should not cut any off at this point.

Bettie: NIH/DOE Comm on Mouse established. They will provide guidance on mouse research. Doing genetic map 1st, physical subsequent.

Berg: Mouse people should be w humans

Olson: if we're going to add goals we could meet - genetic map of mouse is very achievable.

Informatics

D. Benson: goal #1 - we are there now

E. Branscomb: goal #2 - database tools - good progress; being studied at Los Alamos.

Berg - ? poses - what genes have been cloned on the X chromosome? - unanswerable at this time.

Branscomb: ? may be overall, turnkey databases to open in necessary ways.

Olson – shouldn't gDB be able to support this?

DBenson – yes

Cox – anything published will be there. Will you give you a GenBank # for sequencing

goal #3 – DB: it is done, it will be done, will never be finished. So is an open goal.

EJ – Depends on how many seq. we have. Diff kind of goal than the others.

ELSI:

Walker: For the companies; the push is for diagnostics. This pushes ELSI issues. With CF there is a drive to have medical community use tests. But they lack stratification/training to use these tests. Access also a question.

Develop a program – gone a long way. NIH + DOE have issued PA – DOE is pursuing tools based on privacy. NIH – RFA on CF → 32 applics, some good scores. NIH will do some clinical costs this study → good data on preliminary?

ELSI NIH program includes 6 on education.
NIH genome office did videotape - distributed broadly.

ELSI Wkng gp determined that answers would not come fast enough. Set 4 high 10 areas:
① delivery of gen services - CF
② 10m project of del/access/distrib of genetic services
③ privacy - DOE wtg last July → Bethesda wtg next week. To brief Dave Galas / Watson before Congressional hearings.

Part of challenge for ELSI - how specific should we be about genetic info - same issues for medical info in general - we have been a catalyst for these areas. Discussion in Working gp if it should be state-by-state or national policy. This is ongoing debate.
③ insurance - Task Force wtg in December.
Developing actuarial data
④ discrimination - esp in employment. Last ASC wtg discussed ways in which EEOC was affecting protections of ADA - not protecting on genotype, testing not only on those things that affect job performance, and privacy.
Berg/Wolf letter to EEOC. EEOC letter said little in bureaucratic lang. We'll seek interpretation with disabilities savvy people. 3 kinds of disabil: ① disabled ② history of disability or ③ appearance of disability

Example of #3 - Turn victim who might look disabled but able to do job. We need to have genetic protection under disabilities. EEOC: "there are certain circumstances ... asymptomatic ... genetic diseases" Late onset gen diseases people should be protected - useful to have.

Berg - not asymptomatic - only when it is expressed.

WEXLER - will seek clarification of whether this was intended to be language of the 3rd definition.

to WEXLER Task Force
Berg - do you screen yourself to unrealistic expectations? What if we offer no medical hope for genetic diseases? How do we tone down rhetoric - education on allied disease genes.

Watson - we are identifying the genes only. We do not have the $ to develop the therapy.

Berg - we need to be clearer with this message. McKusick diseases where do we go from knowing the gene? People are going to see a lot of scary information.

Watson - good part will be that 90 % of all ♀ will determine that they don't have the breast cancer gene.

Wexler: trying to address the issues in a # of ways - need to be realistic about HGP, be direct about educational materials and the promise of the HGP, access to services - providers are not getting necessary information Must couple info c̄ counseling.

Berg - these are also open ended goals

Watson - clear that budget for Ethics increase 3 → 5%. Will be 10% in next few years. Area of Congressional interest. If we find these ca genes, our $ will be inadequate to address issues.

Caskey - need to ensure standards here in keeping to medical practice standards. Should not hype this out to the point of paralysis. Need to mix in other medical specialties in decision making. CF test = 92% accuracy - very good compared to other fields.

Zinder: ELSI program holds medical practice to the highest standards as it should - and make recommendations.

Caskey - just make the standard reasonable.

Cox: HGP was only place in country to peer & where mouse is – should be proud of CF efforts. Stirred NIH to action.

Wexler – 1 more area – large studies of hundreds – how to publish data – who to tell. Eric dev. guidelines

Training

Bettie: 4 predoc trng awards → 30 students in 1st yr
awards to 6 minority students
20 postdocs – most have degrees in molec bio
not as successful in recruiting mathematicians + computer specialists → Special Emphasis Research Career Award established (SERCA). Hoping this will be attractive + accomplish cross training. This Announcement will be out by end of Sept.
Courses: to get technology into more labs. 1st Ann → 6 awards. Will reannounce as PA – ongoing announcement.

EJ: this goal has already been modified. #s not so important, but are looking at types of people.

Berg – do Ctrs have training

J Peterson – some, not much. A graduate student +

Post-doc on each.

Bettie - training program is broader than just research centers

Berg - we train in our dept - just matter of $.

Watson - we have been disappointed by # of applications.

Bettie - perception in community that NIH does not have training $ so HSP does not either. Working to correct that.

Pearson - do we need to modify this spent $ for training into less core - Change emphasis. Better training by doing than by specialized courses.

Hood - at Caltech - training is working quite well. Drexel is having a good biological sciences home - cannot be in engineering. Need to work at it. Need to think about how to bring in people we need.

Bettie - 1st award to Fellow at Affymax

DOE training rpt - Galas
10 post docs this year: lge # cross disciplinary. Many students in Ctrs are molec biologists. Critical to bring in experts in other fields or vice versa. The Centers are the places where effective training can take place.

Watson: ? to Cox - what of training programs aimed at genetic diagnosis

Cox: Would be useful. Have to have interest in genomics. Disagree a bit about Centers - much of creativity occurs in indiv labs - true in all disciplines. ? value of putting training slots into Centers.

Watson - if could promise medical school training this would play well in Congress. Other NIH Institutes are avoiding these issues. We need to rethink our training program to train people in diagnostic tests & interpretation.

Hood - also suggest encourage grants from institutions where can mix up the disciplines - encourage real interdisciplinary training programs.

Watson - need to focus on medical schools now.

Technology Development
open ended / motherhood.

Tech transfer
includes patent issue - S Stransberg - will be on Tuesday Head program.

Tech Development
Stransberg - are there other things we should be doing to foster this?

Berg - many things get turned down in this area.

Watson - issue of quality. Study sections are very conservative.

Stransberg - R21 grants may help some. Tech Dev is high risk / challenges.

Lerman - how do you deal w tech dev from academics?

EJ: most grants we fund are from academics.

[illegible]: trying to encourage industry.

Guyer: R21 was to get around absence of preliminary data

Hood - 2 kinds of tenders
① ventures into known area - RO1s
② centers / interdisciplinary areas.
need to assess both areas

J. Peterson: some examples of good tenders at Centers - Esp Wash U and Michigan.

Guyer: one measure is interaction of ctrs & industry - ct Beckman

JP: during P20 developmental grant being introduced for interdisciplinary work - tender is apart of that initiative.

Branscomb - tender terms do have a grantsmanship problem. - need to work w them.

Diane Smith - if pages are labeled proprietary then under FOI commercial rights are not violated.

Weiler - ELSI wrote a guide to writing a grant - could share.

Pearson - this is a hot button for industry - confidentiality for grants/legal issues. Need to be clearer to industrial people who do not know how game is played. Thought they were confidential already.

D. Smith - difference between confidentiality & proprietary.

Pearson - we need to spell these essies out very clearly to these people.

Technology Transfer

Berg - nothing to add; a platitude so easy to achieve.

PAC mtg - Jan - Informatics

Watson - minute index markers people
MG - do in June before mtg - day before.

Galas - goals of Plan should stay as they are? with reduced funding? [Remain on a year]
Berg - too easily to modify.

International Program.

David Hinson = Hugo rpt - Changes:

① Wyngaarden resigned - member of Council/ active input to US office. Actively searching for replacement. Need to have active managerial direction. 18 member Council meets 2x/yr.

② US office - enormous changes - Am office in Bethesda part of HHMI. Sept 1 payment from HHMI stops. Grant from HHMI of $1m/4 yrs. Actively negotiating w Johns Hopkins to provide the infrastructure needed. Documents between attorneys at this point.

③ European office in London - backed by Bodmer organization = prototype for US program: Bodmer provides salaries & expenses. European office is better established organizationally.

④ Japan office - genome office under change. Have promised support for Hugo. Hope to see some consolidation so work only w 1 group

⑤ Moscow office - set up July 1 as satellite office of European office.

⑥ Brunenhoder + Dane will meet w NATO in Brussels in October.

Mtg in London was critical one - Council members made a real commitment. Purpose of organization still not clear; do not control purse strings; work w many agencies - not always easy to walk fine line.

? if Hugo should continue.

Wolf: Any coord between Hugo + Japan activities

Diane: Japanese are unique – not run like anybody else. Science + funding have been on 2 different tracks. They want to become a major player + work w the US so there may be changes in this area. Japanese are fragmented more much infighting.

Jasper – what do you do?

Diane: gp of scientists worried about govts making decisions – make recommendations on funding. Has become something different – longer + organizing programs. Original purpose was to coordinate. Doing chromosome-specific workshops. Have cosponsored scientific mtgs. Launching an ELSI program internally. Also – intellectual property rights. – second mtg in Europe in fall.

Cantor – Hugo began as elitist Academy model. Character has really changed. Japanese problem – have associated w many people in many ways. STA is program to deal with. Science + funding are not related.

When Hugo started - Human gene mapping Workshops were conducted. Program grew. Hugo committee took over this program. Financially impossible to continue mtgs in same style. Most recent mtg in London cost $2m pounds. Hugo promoting single chromosome workshops. This will require much coordination. Have an annual mtg to share issues. Problem for groups that relate to multiple chromosomes - cannot run from room to room. Proposals - attach a mtg to coordinated mtgs or to Human genome mtgs (San Diego). The mtgs have been useful but have become obsolescent.

Hood - Hugo has been a failure because of this fn. Until we disengage from this, Hugo will not WORK. There are many useful ways Hugo could get involved.

Cox: international workshops were valuable. Were done in by marrying data input w mtg. Do not need to have computers there at all. go to interact w scientists, not input data. Very mixed bag of participants. Real need for cross-chromosome meetings.

WEXLER - how separate functions?

Cantor - Hugo does not have resources to do all single chromosome workshops.

At London mtg - had workshop on CDNAs + YACs. No active gp knew what another group was doing - total info vacuum. Very different US/European policies.

Hugo has spawned 2 small gp's to collect info on CDNAs + YACs - reports this fall.

CDNAs: major US gps are seeking patents - then make available. In Europe - no info made available yet. Will be available to subscribers perhaps. This is EC + UK policy made by MRC. It will be free to academics once a company policy is established. European patent attorneys say they are not patentable.

Hugo plans ① to collect this info, list programs + contacts + procedures.

② search databases to see if peg already there + who did it. Technically feasible to do this and keep privacy intact.

③ exchange characterized CDNAs to avoid redundancy. One caveat - no Japanese rep at mtg - not yet clear what they will do.

US more interested in high resolution maps than Europeans.

Major gp - none knew what anyone else was doing.

Galas: ? of access - might as well repeat research if cannot access them.

Candor - hopes this is a temporary problem.

Wasson - GDB - interest in international funding. Wasson/Galas letter encouraged international funding. Mtg in London w Japanese + EC -> mtg Dec 16 in Paris to discuss officially. Japanese prepared to fund fight away. Formula - 40% US, 40% EC, 20% Japanese. We have only funded 1st yr - seeking agreement for year 2 ->. 2 advisory gps - ① govt - $ ② scientific - what it does + how much funding should be provided.

negotiations will be painful + slow but it will come. Hugo needs to show it will do something before the issue of international funding can be raised.

Galas: Hugo is the scientists - can only play a role in scientific advisory - not the coordinating comm. Wasson does not want Hugo in science - Galas does not want Hugo in management.

Ef - Hugo + GDB are separate.

Wasson - raised ?s of Hugo future + Bodmer role

Pearson – alternative was Mesco – not very attractive need to try to make Hugo succeed. To restart it would be very difficult. Need to separate Hugo from GDB.

Watson went on

Guyer – interest in having Hugo coord workshops – will establish the organization & make it legitimate. Can then expand.

Cantor – problems can be solved if good Exec Dir who provides real direction.

Cox: this is not countries – this is individuals acting in bad faith.

Cantor – Continue with YACS – entire YSC libraries are being distributed w few strings. In England – not the same – not the whole library. Must guarantee giving information back. Very different situation – Europe has a central repository that users must subscribe to = CEPH model. It is very cumbersome, nothing innovative.

Hugo is collecting info on libraries – what available & how access.

Agreement - valuable to have a common YAC library w probes in public domain. No point in screening probes - - anyone can get a YAC that has been hit with a public domain probe. Goal is to have GDB be the repository of that info.

Also agreed that characterized YACs should be shared as quickly as possible.

Groups are now aware of the differences and can begin to address them.

Watson - need to help our Russian friends. Anxious to ↑ Soviet scientists in US. An mtg in November need to try to establish formal program for this.

Ej: FIC has a program [Watson - want more]

Galas: Cantor, Diane + he visited Soviet Union - very impressive enthusiasm. Need to decide how NIH/DOE should arrange this.

Watson - wants formal program for Public relations. Could fund them increased funds that will be made available.

Galas - endorsed idea. 1 small caution - do not know fate of Academy of Scientists.

Watson - also want to propose an Honorary Degree for contributors to Program - French - acknowledge the role of CEPH. Some ceremony in the US. (Guyer) June mtg on linkage mapping - could be done then.

Bettie - could name a fellowship program for him

Pearson - return to Soviet problem - need to do post doc program quickly - both NIH & DOE could work with them - need to be expansive in thinking.

Cantor: very small amts of currency can make a huge difference in Soviet program.

Watson - begin with US collaborator w Soviets.

Mozzis - people already working with this - Some outstanding Russian scientists - need help.

Galas - real opportunities.

Helen

The enclosed brief abstracts were prepared to complement a presentation at the DOE/NIH Advisors Retreat, on Technology Development facets of the DOE Human Genome. You may wish to review them enroute to the meeting.

LEADING TECHNOLOGY DEVELOPMENT PROJECTS IN THE DOE HUMAN GENOME PROGRAM

Mapping support

New Strategies For The Closure Of The Chromosome 19 Contig Map. Pieter J. de Jong, C. Amemiya, C. Aslandidis, J. Tang, K. Yokobata, and A.V. Carrano. LLNL
Distinctive fingerprints of human sub-genomic elements are rapidly obtained by inter-Alu-PCR. They serve as a uniting language to assign the diverse types of recombinant clones to human genomic regions defined by hybrid cells; bridge contig gaps by a multiplexed cross-correlation of arrayed cosmid and YAC clones; and check validity of contig assignments.

DNA Sequence Mapping By Fluoresence In Situ Hybridization (FISH). Barbara Jo Trask, B. Brandriff, K. Tynan, Ger van den Engh, and A. V. Carrano. LLNL
Chromosome 19 map construction is being facilitated by FISH for localizing cosmids on Ch19; confirming contig end clone assignments; choosing a reference clone set for multicolor FISH on Ch19; ordering and orienting contigs by multicolor FISH; and aiding choices of contig gap closure strategies, by estimating inter-contig distances to 50 kb resolution in interphase nuclei and sperm pronuclei.

Automated Methods for Large-Scale Physical Mapping
Patricia A. Medvick and Robert Hollen. LANL
Developments are concentrated on the identification of petri plate colonies, with an imaging system that directs a robotic arm to transfer colonies to 96 well trays; gridding dense arrays on nylon hybridization membranes from microtiter-well plates; and database development for robotic control and for initial storage of hybridization results.

Overcoming Genome Mapping Bottlenecks.
Charles R. Cantor. LBL
Top down mapping capacities are being improved by extensing PFG techniques toward the human chromosome size range; covalent attaching engineered strepavidin to beads and matrices, thus improving sequence specific DNA capture reagents systems using biotinylated probes; using the capture systems to align and straighten target DNAs for parallel microdissections.

Sequencing strategies

Develop Fully Integrated Technology For Sequencing Genomes As Large As That Of Humans. George M. Church. Harvard University
A multiplex sequencing clone pooling system using Maxam-Gilbert chemistry will be further developed through directed sequencing strategies supporting oligonucleotide walking and hybridization selection methods; computer assisted film reading and base calling with accuracy assignments; and adaptive software accommodating technical advances.

Advanced Sequencing Technology
Raymond F. Gesteland and Robert Weiss. University of Utah
A multiplex sequencing clone pooling system using Sanger biochemistry will be further developed through implementation of directed sequencing from primers on transposons; fractionations through capillary gels; and a base calling algorithm using communication and signal processing theory.

Large-Scale DNA Sequencing with a Primer Library
F. William Studier and John J. Dunn. BNL
The strategy is to sequence 40-kb clones (such as cosmids) directly by primer walking, using only primers from a library. If successful, this strategy would improve efficiency and reduce costs by a least an order of magnitude over current practice, and would provide the basis for developing automated sequencing machines capable of generating perhaps 100,000 bases per hour.

Oligonucleotide Libraries for High-Throughput DNA Sequencing. Kenneth Beattie Genosys Biotechnologies, Inc. (SBIR Award).
A library of 3,314 nonamer oligonucleotides is being synthesized, to support the high throughput genome sequencing by a progressive primer walking strategy of W. Studier (BNL).

Megabase Sequencing of Human Immune Receptor Loci
Leroy E. Hood, California Institute of Technology
Complete sequencing of the three T-cell and receptor receptor and HLA loci in man is progressing. The major goals are to evaluate new strategies, techniques and instrumentation for large-scale DNA sequencing; to determine the sequence for T-cell receptor and HLA loci from humans; and to create new approaches to the study the molecular biology of the immune response with this information.

Sequencing by Hybridization (SBH)
Radomir Crkvenjakov and Radoje Dramanac. ANL
Reliable procedures for interrogation of DNA clones with oligomer probes have been achieved. With dot blots of clones representing a multiple covering of the subject chromosome, the interrogation cycles progressively achieve clone ordering, sequence motif recognition (partial SBH) and complete SBH. Rapid characterization of cDNA libraries can also be achieved by partial SBH.

DNA Sequence Analysis by Solid-Phase Hybridization. Robert S. Foote, R.A. Sachleben, K.B. Jacobson, T.V. Mitchell and R.J. Mural. ORNL
The goal is the synthesis of ordered, two-dimensional arrays of oligonucleotide sequences on planar solid supports, for use as DNA hybridization probes against clone fragments. The target DNA will carry labels that can be detected by optical or mass spectral characteristics. For SBH analyses of cosmids and YACs, arrays with 8-mers to 11-mers are needed.

Automation of the Front End of DNA Sequencing
David Mead and Lloyd Smith. University of Wisconsin, Madison
An automated front-end system for large-scale mapping and DNA sequencing will support solid-phase purification of clonal DNAs and their inserts; simple and efficient DNA fragmentation and fractionation; automatic recloning; microsequencing instrumentation; and a directed end-sequencing strategy. The methodology will include a microliter liquid-dispensing robot.

Fractionation with fluorescense detection systems

The Separation of DNA by Pulsed Field Capillary Electrophoresis.
Barry L. Karger Northeastern University
Capillary gel and pulse field electrophoretic procedures are being combined and optimized, to improve fractionations of double stranded DNA fragments in the 1-50 kb and 50-500 kb size ranges. Effects of wall coatings as well as gel composition are being ascertained. Both restriction mapping and sequencing tasks will be speeded by these improvements.

A High-Speed Automated DNA Sequencer
Lloyd M. Smith University of Wisconsin
The goal of this research is the development of a high-speed automated instrument capable of sequencing 500 bases per hour within 50 lanes of a slab gel. Sequencing ladders form wthin an ultrathin polyacrylamide gel with laser excited fluorescence readout by a CCD. The theoretical daily throughput per instrument is 500 X 50 X 24 = 600,000 bases of raw sequence data per day.

Multi-Column Gel Electrophoresis and Laser-Induced Fluorescence Detection for DNA Sequencing at 64,000 Bases/Hour. Norman J. Dovichi. University of Alberta
A laser-induced fluorescence detector is under development, for concurrently reading out multiple capillaries in which gel electrophoretic is progressing. By reading 32 capillaries simultaneously, each displaying 2000 bands/hr, sequencing rates of 64,000 bases/hr may be achieved with this system. With only one power supply and laser required per system, costs are modest.

Ultrasensitive Fluorescence Detection during DNA Fractionation
Richard A. Mathies and A. N. Glazer. University of California, Berkeley.
Ultrasensitive fluorescence systems being refined including microscope slide sized fractionation gel system with a confocal laser excitation system; exciting DNA-fluor complexes, which are stable during fractionation; and with detection on a stage-scanned confocal fluorescent microscope system. This system will miniaturize all DNA fragment analyses including sequencing.

New labels and companion instrumentation

Development of an Ultrasensitive Detection System for DNA Sequencing. James F. Hainfeld Nanoprobes, Inc. (SBIR).
A new gold particle has been developed that is water soluble, can be reacted with nucleic acids in a specific way, and is highly sensitive to silver developer. This system has the experimentally demonstrated potential to detect 10-20 mol of target. The performance of the this environmentally benign gold particle system is being optimized for hybridization assays.

DNA Sequencing Using Stable Isotopes
K.B. Jacobson, H.F. Arlinghaus , G.M. Brown, R.S. Foote, F.W. Larimer, R.A. Sachleben, N. Thonnard, and R.P. Woychik. ORNL
Stable isotopes of iron, tin and several lanthanides can be used as co-resident labels for nucleic acids. A gel, blot or array can be scanned by a laser or atomic beam to release the isotopic labels for quantitation by RIS (resonance ion spectroscopy). This multiplexing and high scanning rates promise high rates for sequencing and other nucleic acid analyses.

Development of Micron to Sub-Micron Thickness Electrophoresis Gels to Optimize Resolution in DNA Sequencing Using RIS. Norbert Thonnard. Atom Sciences (SBIR)
High speed RIS can be used to measure subattomole quantities of isotopically labeled DNA with excellent lateral and mass resolution of isotopic labels. Only the labels in the first few molecular layers of the gel are detected. Thus micro-thin electrophoresis gels will have sufficient DNA loading, and with their reduced ohmic heating, will allow much higher fractionation speeds.

Ultrasensitive Detection of Luminescence from Lanthanide Ions as Labels for DNA Mapping and Sequencing. Gilbert M. Brown, R. S. Foote, K. B. Jacobson, F. W. Larimer, J. M. Ramsey, R. S. Ramsy and W. B. Whitten. ORNL
When a lanthanide is caged by a UV absorbing moiety, the resultant complex has a high quantum efficiency for secondary luminescence. This radiation is comprised of sharp lines as contrasted to the broad band emission of organic fluors. Thus many co-resident lanthanide labels can be distinguished and will support multiplex labeling protocols for both mapping and sequencing analyses.

Mass spectroscopy fractionation systems

The common promising feature of all the mass spectroscopy systems is that, as contrasted with gel electrophoretic fractionation ssytems, the fractionation time is negligible. Sample preparation will be a time limiting factor when these systems pass proof-of-concept trials.

High-Speed DNA Sequence Analysis by Matrix-Assisted Laser Desorption Mass Spectrometry. L. Smith (U Wisconsin) and B. Chait (Rockfeller University)
A matrix containing products of the Sanger or Maxam-Gilbert sequencing reactions will be vaporized by matrix assisted UV laser desorption. A resolved mass spectrum of a mixture containing DNA fragments up to 500 bases in length is sought. An automatical comparison of the mass spectra from the set of four sequencing reactions will generate the finished sequence.

Vacuum Ultraviolet (VUV) Ionizer Mass Spectrometer for Genome Sequencing. Winston H. Chen, M.G. Payne and K.B. Jacobson. ORNL
Sanger sequencing fragments will have an adduct which can be ionized by VUV, providing for mass spectroscopic analysis of fragments without other chain breakage. The four Sanger products will be individually ionized and data combined, with projected sequence reads of upto 3000 bases. Construction of the mass spectrometer equipped with the VUV photoionizer is in progress.

Time of Flight Mass Spectrometry of DNA for Rapid Sequence Determination Peter Williams and N. Woodbury. Arizona State University
Sanger reaction products will ablated from ice by laser pulses. The ablated DNA molecules are ionized by attachment of sodium ions, produced by resonant multiphoton ionization of sodium atoms in the vaporized solution. Intact DNAs up to 9 kb have been volatilized, and the projected resolution of fragments by the subsequent time-of-flight mass spectroscope is in the 450 base range.

Single Molecule Analyses

High-Resolution DNA Mapping by Scanning Transmission Electron Microscopy (STEM). James F. Hainfeld BNL.
Undecagold clusters have been developed for very high contrast electromicroscopic applications. When used as a lable on a DNA probe, by D-loop hybridization to duplex DNAs, positional localizations to 3-5 bp are achieved. This capacity far surpasses that attainable by FISH. STEM systems generally support very high data acquisition rates.

Scanning Molecular Exciton Microscopy A New Approach To Gene Sequencing. Raoul Kopelman, J. Langmore, V. Makarov, and B. Orr. University of Michigan
In Molecular Excition Microscopy (MEM) excitons of energy are exchanged between a scanning tip and the target. Exchange is modulated by the detailed orbital structure of the target. For application to DNA sequencing, the four bases will be labeled with distinct metal ions to achieve the contrast necessary for distinguishing the bases.

Scanning Tunneling Microscopy
T.L. Ferrell, R.J. Warmack, and D.P. Allison. ORNL
A scanning tunneling microscope with single-atom resolution is used to image atomic structure on surfaces, to alter atomic positions, and to probe the dynamical phenomena caused by collective electron motion and motion of ions. Development includes extending capabilities to a wider range of materials, to identification of atomic species, and to studies of biological samples.

Rapid DNA Sequencing Based Upon Fluorescence Detection Of Single Molecules
J.H. Jett, R.A. Keller, J.C. Martin, and E.B. Shera. LANL
A technique is being developed for the rapid sequencing of 40-kb fragments of DNA at rates up to several hundred bases per second. The four bases in a single fragment of DNA have distinguishing labels, the labeled fragment is in a sample stream, and individual bases are recognized by their fluorescence as they are processively cleaved off and pass through the laser beam.

Single Molecule Detection of Nucleotides Tagged with Fluorescent Dyes. S.A. Soper, L.M. Davis, J.H. Jett, R.A. Keller, J.C. Martin, E.B. Shera. LLNL
Candidates for single molecule detection are being characterized for absorption cross section, fluorescence quantum efficiency, fluorescence lifetime and the photodestruction efficiency. Single molecule detection has been achieved for some candidate flurescent dyes attached to nucleotides using the pulsed-laser single molecule detection apparatus.

Single Molecule Detection Using Charge-Coupled Device Array Technology. Bonner Denton. University of Arizona
A Charge-Coupled Device (CCD) is being developed to complement the LANL system, with a special operating mode to enhance the discrimination between fluorescence from a single molecule and the background radiation. Register shifts between rows in the CCD will match the flow velocity of released bases, so that fluorescence from a single molecule will be collected in a single moving charge packet with an area approaching a single pixel.

Informatics

BISP: VLSI Solutions to Sequence Comparison Problems
Tim Hunkapiller, L. Hood (Cal Tech), E. Chen (JPL), and M. Waterman (USC)
BISP represents a systolic hardware implementation of a dynamic programming algorithm of Smith and Waterman, optimizing its ability to define local similarities. It is optimized for the determination of any number of local similarities between pairs of sequences, returns values that will allow for the reconstruction of the alignment; and is specifically designed to employ complex, user-definable similarity rules.

Prepared by Daniel Drell and Marvin Stodolsky, DOE

LEADING TECHNOLOGY DEVELOPMENT PROJECTS IN THE DOE HUMAN GENOME PROGRAM

Mapping support

New Strategies For The Closure Of The Chromosome 19 Contig Map. Pieter J. de Jong, C. Amemiya, C. Aslandidis, J. Tang, K. Yokobata, and A.V. Carrano. LLNL
Distinctive fingerprints of human sub-genomic elements are rapidly obtained by inter-Alu-PCR. They serve as a uniting language to assign the diverse types of recombinant clones to human genomic regions defined by hybrid cells; bridge contig gaps by a multiplexed cross-correlation of arrayed cosmid and YAC clones; and check validity of contig assignments.

DNA Sequence Mapping By Fluoresence In Situ Hybridization (FISH). Barbara Jo Trask, B. Brandriff, K. Tynan, Ger van den Engh, and A. V. Carrano. LLNL
Chromosome 19 map construction is being facilitated by FISH for localizing cosmids on Ch19; confirming contig end clone assignments; choosing a reference clone set for multicolor FISH on Ch19; ordering and orienting contigs by multicolor FISH; and aiding choices of contig gap closure strategies, by estimating inter-contig distances to 50 kb resolution in interphase nuclei and sperm pronuclei.

Automated Methods for Large-Scale Physical Mapping
Patricia A. Hedvick and Robert Hollen. LANL
Developments are concentrated on the identification of petri plate colonies, with an imaging system that directs a robotic arm to transfer colonies to 96 well trays; gridding dense arrays on nylon hybridization membranes from microtiter-well plates; and database development for robotic control and for initial storage of hybridization results.

Overcoming Genome Mapping Bottlenecks.
Charles R. Cantor. LBL
Top down mapping capacities are being improved by extensing PFG techniques toward the human chromosome size range; covalent attaching engineered strepavidin to beads and matrices, thus improving sequence specific DNA capture reagents systems using biotinylated probes; using the capture systems to align and straighten target DNAs for parallel microdissections.

Automation of Hybridization Analysis. J. Jaklevic, T. Hanson, W. Kolbe, E. Theil, D. Uber. LBL
Previous work in developing an automatic colony picking and imaging acquisition system is being extended to an automated analysis of hybridizations. It permits imaging, digitization and localization of each positive signal in an array, as well as assignment of confidence levels. Data then becomes part of a general laboratory information management system.

Sequencing strategies

Develop Fully Integrated Technology For Sequencing Genomes As Large As That Of Humans. George M. Church. Harvard University
A multiplex sequencing clone pooling system using Maxam-Gilbert chemistry will be further developed through directed sequencing strategies supporting oligonucleotide walking and hybridization selection methods; computer assisted film reading and base calling with accuracy assignments; and adaptive software accommodating technical advances.

Advanced Sequencing Technology
Raymond F. Gesteland and Robert Weiss. University of Utah
A multiplex sequencing clone pooling system using Sanger biochemistry will be further developed through implementation of directed sequencing from primers on transposons; fractionations through capillary gels; and a base calling algorithm using communication and signal processing theory.

Large-Scale DNA Sequencing with a Primer Library
F. William Studier and John J. Dunn. BNL
The strategy is to sequence 40-kb clones (such as cosmids) directly by primer walking, using only primers from a library. If successful, this strategy would improve efficiency and reduce costs by a least an order of magnitude over current practice, and would provide the basis for developing automated sequencing machines capable of generating perhaps 100,000 bases per hour.

Oligonucleotide Libraries for High-Throughput DNA Sequencing. Kenneth Beattie Genosys Biotechnologies, Inc. (SBIR Award).
A library of 3,314 nonamer oligonucleotides is being synthesized, to support the high throughput genome sequencing by a progressive primer walking strategy of W. Studier (BNL).

Megabase Sequencing of Human Immune Receptor Loci
Leroy E. Hood, California Institute of Technology
Complete sequencing of the three T-cell and receptor receptor and HLA loci in man is progressing. The major goals are to evaluate new strategies, techniques and instrumentation for large-scale DNA sequencing; to determine the sequence for T-cell receptor and HLA loci from humans; and to create new approaches to the study the molecular biology of the immune response with this information.

Sequencing by Hybridization (SBH)
Radomir Crkvenjakov and Radoje Dramanac. ANL
Reliable procedures for interrogation of DNA clones with oligomer probes have been achieved. With dot blots of clones representing a multiple covering of the subject chromosome, the interrogation cycles progressively achieve clone ordering, sequence motif recognition (partial SBH) and complete SBH. Rapid characterization of cDNA libraries can also be achieved by partial SBH.

DNA Sequence Analysis by Solid-Phase Hybridization. Robert S. Foote, R.A. Sachleben, K.B. Jacobson, T.V. Mitchell and R.J. Mural. ORNL
The goal is the synthesis of ordered, two-dimensional arrays of oligonucleotide sequences on planar solid supports, for use as DNA hybridization probes against clone fragments. The target DNA will carry labels that can be detected by optical or mass spectral characteristics. For SBH analyses of cosmids and YACs, arrays with 8-mers to 11-mers are needed.

Automation of the Front End of DNA Sequencing
David Mead and Lloyd Smith. University of Wisconsin, Madison
An automated front-end system for large-scale mapping and DNA sequencing will support solid-phase purification of clonal DNAs and their inserts; simple and efficient DNA fragmentation and fractionation; automatic recloning; microsequencing instrumentation; and a directed end-sequencing strategy. The methodology will include a microliter liquid-dispensing robot.

Fractionation with fluorescense detection systems

The Separation of DNA by Pulsed Field Capillary Electrophoresis.
Barry L. Karger Northeastern University
Capillary gel and pulse field electrophoretic procedures are being combined and optimized, to improve fractionations of double stranded DNA fragments in the 1-50 kb and 50-500 kb size ranges. Effects of wall coatings as well as gel composition are being ascertained. Both restriction mapping and sequencing tasks will be speeded by these improvements.

A High-Speed Automated DNA Sequencer
Lloyd M. Smith University of Wisconsin
The goal of this research is the development of a high-speed automated instrument capable of sequencing 500 bases per hour within 50 lanes of a slab gel. Sequencing ladders form wthin an ultrathin polyacrylamide gel with laser excited fluorescence readout by a CCD. The theoretical daily throughput per instrument is 500 X 50 X 24 = 600,000 bases of raw sequence data per day.

Multi-Column Gel Electrophoresis and Laser-Induced Fluorescence Detection for DNA Sequencing at 64,000 Bases/Hour. Norman J. Dovichi. University of Alberta
A laser-induced fluorescence detector is under development, for concurrently reading out multiple capillaries in which gel electrophoretic is progressing. By reading 32 capillaries simultaneously, each displaying 2000 bands/hr, sequencing rates of 64,000 bases/hr may be achieved with this system. With only one power supply and laser required per system, costs are modest.

Ultrasensitive Fluorescence Detection during DNA Fractionation
Richard A. Mathies and A. N. Glazer. University of California, Berkeley.
Ultrasensitive fluorescence systems being refined including microscope slide sized fractionation gel system with a confocal laser excitation system; exciting DNA-fluor complexes, which are stable during fractionation; and with detection on a stage-scanned confocal fluorescent microscope system. This system will miniaturize all DNA fragment analyses including sequencing.

New labels and companion instrumentation

Development of an Ultrasensitive Detection System for DNA Sequencing. James F. Hainfeld Nanoprobes, Inc. (SBIR).
A new gold particle has been developed that is water soluble, can be reacted with nucleic acids in a specific way, and is highly sensitive to silver developer. This system has the experimentally demonstrated potential to detect 10-20 mol of target. The performance of the this environmentally benign gold particle system is being optimized for hybridization assays.

DNA Sequencing Using Stable Isotopes
K.B. Jacobson, H.F. Arlinghaus , G.M. Brown, R.S. Foote, F.W. Larimer, R.A. Sachleben, N. Thonnard, and R.P. Woychik. ORNL
Stable isotopes of iron, tin and several lanthanides can be used as co-resident labels for nucleic acids. A gel, blot or array can be scanned by a laser or atomic beam to release the isotopic labels for quantitation by RIS (resonance ion spectroscopy). This multiplexing and high scanning rates promise high rates for sequencing and other nucleic acid analyses.

Development of Micron to Sub-Micron Thickness Electrophoresis Gels to Optimize Resolution in DNA Sequencing Using RIS. Norbert Thonnard. Atom Sciences (SBIR)
High speed RIS can be used to measure subattomole quantities of isotopically labeled DNA with excellent lateral and mass resolution of isotopic labels. Only the labels in the first few molecular layers of the gel are detected. Thus micro-thin electrophoresis gels will have sufficient DNA loading, and with their reduced ohmic heating, will allow much higher fractionation speeds.

Ultrasensitive Detection of Luminescence from Lanthanide Ions as Labels for DNA Mapping and Sequencing. Gilbert M. Brown, R. S. Foote, K. B. Jacobson, F. W. Larimer, J. M. Ramsey, R. S. Ramsy and W. B. Whitten. ORNL
When a lanthanide is caged by a UV absorbing moiety, the resultant complex has a high quantum efficiency for secondary luminescence. This radiation is comprised of sharp lines as contrasted to the broad band emission of organic fluors. Thus many co-resident lanthanide labels can be distinguished and will support multiplex labeling protocols for both mapping and sequencing analyses.

Mass spectroscopy fractionation systems

The common promising feature of all the mass spectroscopy systems is that, as contrasted with gel electrophoretic fractionation ssytems, the fractionation time is negligible. Sample preparation will be a time limiting factor when these systems pass proof-of-concept trials.

High-Speed DNA Sequence Analysis by Matrix-Assisted Laser Desorption Mass Spectrometry. L. Smith (U Wisconsin) and B. Chait (Rockfeller University)
A matrix containing products of the Sanger or Maxam-Gilbert sequencing reactions will be vaporized by matrix assisted UV laser desorption. A resolved mass spectrum of a mixture containing DNA fragments up to 500 bases in length is sought. An automatical comparison of the mass spectra from the set of four sequencing reactions will generate the finished sequence.

Advanced Detectors for Mass Spectrometry. J.M. Jaklevic, W.H. Benner and J.E. Katz. LBL
A major limitation to the extension of mass spectrometry to a molecular size region where direct sequencing becomes practical involves the decrease in sensitivity inherent in existing methods for ion detection. We propose to investigate these limitations using an existing particle generation system and develop improved methods for ion detection based on several feasible alternative technologies. A time-of-flight mass spectrometer incorporating detector and electronics improvements will be designed and made available for research applications involving direct sizing of DNA molecules in sequencing and mapping applications.

Vacuum Ultraviolet (VUV) Ionizer Mass Spectrometer for Genome Sequencing. Winston H. Chen, M.G. Payne and K.B. Jacobson. ORNL
Sanger sequencing fragments will have an adduct which can be ionized by VUV, providing for mass spectroscopic analysis of fragments without other chain breakage. The four Sanger products will be individually ionized and data combined, with projected sequence reads of upto 3000 bases. Construction of the mass spectrometer equipped with the VUV photoionizer is in progress.

Time of Flight Mass Spectrometry of DNA for Rapid Sequence Determination Peter Williams and M. Woodbury. Arizona State University
Sanger reaction products will ablated from ice by laser pulses. The ablated DNA molecules are ionized by attachment of sodium ions, produced by resonant multiphoton ionization of sodium atoms in the vaporized solution. Intact DNAs up to 9 kb have been volatilized, and the projected resolution of fragments by the subsequent time-of-flight mass spectroscope is in the 450 base range.

Single Molecule Analyses

High-Resolution DNA Mapping by Scanning Transmission Electron Microscopy (STEM). James F. Hainfeld BNL.
Undecagold clusters have been developed for very high contrast electromicroscopic applications. When used as a lable on a DNA probe, by D-loop hybridization to duplex DNAs, positional localizations to 3-5 bp are achieved. This capacity far surpasses that attainable by FISH. STEM systems generally support very high data acquisition rates.

Scanning Molecular Exciton Microscopy A New Approach To Gene Sequencing. Raoul Kopelman, J. Langmore, V. Makarov, and B. Orr. University of Michigan
In Molecular Excition Microscopy (MEM) excitons of energy are exchanged between a scanning tip and the target. Exchange is modulated by the detailed orbital structure of the target. For application to DNA sequencing, the four bases will be labeled with distinct metal ions to achieve the contrast necessary for distinguishing the bases.

Imaging of Biological Structures using Atomic Force Microscopy. William F. Kolbe and Miguel B. Salmeron. LBL
Biological structures and molecules including DNA are being studied with atomic force microscopy (AFM) with applications directed towards molecular manipulation, mapping and sequencing. Key components of this research include the design of appropriate instrumentation and the development of techniques for the substrate attachment and manipulation of the DNA molecules prior to imaging.

Scanning Tunneling Microscopy
T.L. Ferrell, R.J. Warmack, and D.P. Allison. ORNL
A scanning tunneling microscope with single-atom resolution is used to image atomic structure on surfaces, to alter atomic positions, and to probe the dynamical phenomena caused by collective electron motion and motion of ions. Development includes extending capabilities to a wider range of materials, to identification of atomic species, and to studies of biological samples.

Rapid DNA Sequencing Based Upon Fluorescence Detection Of Single Molecules
J.H. Jett, R.A. Keller, J.C. Martin, and E.B. Shera. LANL
A technique is being developed for the rapid sequencing of 40-kb fragments of DNA at rates up to several hundred bases per second. The four bases in a single fragment of DNA have distinguishing labels, the labeled fragment is in a sample stream, and individual bases are recognized by their fluorescence as they are processively cleaved off and pass through the laser beam.

Single Molecule Detection of Nucleotides Tagged with Fluorescent Dyes. S.A. Soper, L.M. Davis, J.H. Jett, R.A. Keller, J.C. Martin, E.B. Shera. LLNL
Candidates for single molecule detection are being characterized for absorption cross section, fluorescence quantum efficiency, fluorescence lifetime and the photodestruction efficiency. Single molecule detection has been achieved for some candidate flurescent dyes attached to nucleotides using the pulsed-laser single molecule detection apparatus.

Single Molecule Detection Using Charge-Coupled Device Array Technology. Bonner Denton. University of Arizona
A Charge-Coupled Device (CCD) is being developed to complement the LANL system, with a special operating mode to enhance the discrimination between fluorescence from a single molecule and the background radiation. Register shifts between rows in the CCD will match the flow velocity of released bases, so that fluorescence from a single molecule will be collected in a single moving charge packet with an area approaching a single pixel.

Informatics

BISP: VLSI Solutions to Sequence Comparison Problems
Tim Hunkapiller, L. Hood (Cal Tech), E. Chen (JPL), and M. Waterman (USC)
BISP represents a systolic hardware implementation of a dynamic programming algorithm of Smith and Waterman, optimizing its ability to define local similarities. It is optimized for the determination of any number of local similarities between pairs of sequences, returns values that will allow for the reconstruction of the alignment; and is specifically designed to employ complex, user-definable similarity rules.

Tools for Defining and Manipulating Database Objects. V. H. Harkowitz, A. Shoshani, E. Szeto. LBL
QUEST is a tool being designed to assist users in interactively specifying database queries in terms of objects. It does not require familiarity with query language or operators, but permits selection of items (objects, attributes, etc.) and a guide through the process of specifying queries.