United States Government

Department of Energy

memorandum

DATE May 6, 1986

EPLY TO
ATTN OF ER-70

UBJECT Information on a Major New Initiative: Mapping and Sequencing the Human Genome

TO Alvin W. Trivelpiece, Director
Office of Energy Research

BACKGROUND:

In the early 1970's Walter Gilbert, a former Harvard University physics professor, and Fred Sanger, a Cambridge biochemist, developed important new approaches to DNA sequencing. The methods, which allow rapid determination of the information content of genes, have been a major impetus to the recent burgeoning of academic and commercial biotechnology, and led to Nobel Prizes for Gilbert and Sanger. Because sequencing touches virtually every aspect of modern molecular biology, and bears directly on central questions of heritable mutations, cancer, genetic disease, etc., considerable pressure exists to accelerate the rate at which genes are sequenced and precisely mapped onto chromosomes. At the present rate of sequencing, several hundred years will be required to obtain the full human genome; i.e., to obtain the sequence of every human gene.

In December 1985, I asked Dr. Mark Bitensky, a Los Alamos Senior Fellow to organize a workshop consisting of world leaders in molecular biology and medical genetics, drawn from universities, the private sector and the National Laboratories, to address a number of questions related to stimulating a major increase in the rate of mapping and sequencing human genes. Specifically, participants were asked to assess (i) the feasibility of sequencing the human genome by approximately the year 2000; (ii) the costs of such a venture; (iii) the desirability in terms of human health and national economic growth, and (iv) the nature of the role, if any, that the Department of Energy might play. An executive summary of the meeting, written by Dr. Bitensky, as well as the reports of the participants on which it is based, are attached.

DISCUSSION:

Virtual unanimity was reached on the following conclusions:

(1) Mapping and sequencing the human genome within the next 10-15 years is, under a reasonable set of assumptions, technically feasible, and would be a major achievement in the history of biology.

(2) Although the implications of such an accomplishment cannot be fully anticipated, the project will affect virtually every area of biomedical science, and a substantial number of major medical and basic research advances having important health and economic impacts are expected. For example, we can anticipate important advances in understanding genetic disease; in understanding at the deepest level, the variation in human susceptibility to environmental contaminants and other disease causing agents; in assessing the frequency and nature of heritable mutations; in the design of peptides and protein binding molecules for pharmaceuticals, disease prevention, etc. The participants noted that even a 1% impact on the nation's $400 billion per year health care budget would, within a year, more than return the estimated $1-2 billion cost of the project.

(3) The National Laboratories have a major role to play in a human genome project especially in its initial phases. The role derives in part from the fact that questions central to the OHER mission would become directly addressable; in part from current National Laboratory activities such as the GENBANK and GENE LIBRARY projects at Los Alamos and Livermore, which are the natural precursors of a genome project; and in part from the genome project's crucial dependence on new engineering and computational technologies whose development will require large interdisciplinary teams. The goal of the latter would be to increase sequencing speed by at least two orders of magnitude. This will require robotics development, image processing, physical analysis of separation processes and so forth, as well as various computational requirements including database restructuring, the development of advanced algorithms and analytical capabilities, and national networking with a dedicated supercomputer at the central node.

(4) The project would be national in scope, somewhat reminiscent of the effort that led to the conquest of space, but supported by multiple agencies, including the private sector, with one agency playing the lead, managerial role. International cooperation would be sought at the outset.

(5) The workshop participants were acutely aware of the need for innovations to effectively manage a project that would be far larger and far more interactive than any that has ever been attempted in the life sciences. The extensive experience of the Department of Energy in managing such projects is in distinct contrast to that of other agencies that would also support the genome project. For this, as well as the technical reasons indicated above, DOE is a natural organizational to play the lead management role, at least in the initial phases of the project.

(6) An item for immediate action is the formation of a steering committee to provide guidance on scientific strategies and priorities, administrative structure, and organizational liaison.

Charles DeLisi

Charles DeLisi, Ph.D.
Associate Director for Health
and Environmental Research
Office of Energy Research

Attachments

cc: J. Decker, ER-2

nited States Government

Department of Energy

nemorandum

DATE May 6, 1986

PLY TO TN OF ER-70

SUECT Information on Human Genome Project

TO Alvin W. Trivelpiece, Director
Office of Energy Research

My current thinking is that the overall project might best be divided into two phases. The first phase, to be completed in approximately five or six years, would have three goals: (1) the production of a complete physical map of the human genome; i.e., determination of the chromosomal location of each gene, but not necessarily the sequence of DNA bases; (2) the development of high speed, fully automated methods for sequencing DNA; (3) the development of computerization, as indicated below. Phase II would concentrate on the actual sequencing, and the detailed strategy pursued would be heavily dependent upon the results of the first phase. The attached material, which you may wish to use as a basis for discussion with the Secretary, involves only the first phase, with some emphasis on what we would like to accomplish in the next 18 months.

Charles DeLisi

Charles DeLisi, Ph.D.
Associate Director for Health
and Environmental Research
Office of Energy Research

Attachment

BACKGROUND:

Definition of the Problem. One of the central tasks of modern biology is to determine the physical locations of the genes that determine important human characteristics, such as tendencies toward susceptibility or resistance to certain diseases. Technical advances in the 1970's not only made possible the determination of such genetic maps, but also provided a means for characterizing, in molecular detail, the chemical composition of genes.

Roughly speaking, a gene can be characterized as a long, linear arrangement of four different types of chemical units. These units, or bases, form the letters of an alphabet. Their arrangement, or sequence, can contain information, just as the arrangement of a sequence of letters drawn from the English alphabet can contain information. The scientific knowledge gained by locating and by decoding genes, touches virtually every aspect of modern biology, and bears directly on such issues as the relative importance of environmental and genetic components of disease, the heritability of mutations, and the etiology of cancer. Because molecular studies have such wide-ranging implications, and because they have been a major impetus to the burgeoning of academic and commercial biotechnology, considerable pressure exists to accelerate the rate at which genes are sequenced and precisely mapped onto chromosomes.

The Santa Fe Workshop. On March 3 and 4, 1986, Los Alamos National Laboratory organized a workshop of world leaders in molecular biology and medical genetics--drawn from universities, the National Laboratories and private sector-- to address a number of questions related to stimulating a

major increase in the rate of mapping and sequencing human genes. The charge and conclusions are described in the attached memorandum. Briefly, the participants concluded that major economic and health benefit would accrue from such an effort and that the National Laboratories had a major role to play. The latter conclusion derives in part from the anticipated programmatic impact of the project; in part from its crucial dependence on new engineering and computational technologies whose development will require large interdisciplinary teams, and in part from OHER leadership in closely related work in progress. The participants also noted the need for innovative management of a project that would be far larger and far more interactive than any ever before attempted in the life sciences. The extensive experience of the Department of Energy in managing very large, coordinated, scientific projects, is in distinct contrast to that of other agencies that would also support the genome project, and suggests the project could benefit with DOE in the lead role.

Most of the participants at the Sante Fe workshop were not DOE contractors. The same week, Nobellate Renato Dulbecco, who was not at the meeting, published an editorial in Science urging a national effort to sequence the human genome.

o Obtain the information necessary to determine the physical locations of all genes on chromosomes 16 and 19.

Comment: These two chromosomes are chosen for their relatively small size, their immediate availability, and their medical and scientific importance. There is good reason to believe that knowledge of the detailed maps of these chromosomes will substantially accelerate research progress in a number of important areas including (a) enzymes that repair genes damaged by carcinogens, (b) certain neurological disorders, (c) spontaneous abortions, and (d) genes that cause cancer.

o Begin computer stimulation of mapping and sequencing strategies.

Comment: Because of the enormity of scale for the overall project, the best choice of mapping and sequencing strategy is not obvious. Computer simulations will be used to compare different strategies. This task will have a substantial impact on the long-term direction and strategies of the project.

o Begin computer simulation of data flow and networking.

Comment: The data flow generated by the human genome project will exceed current levels by about three orders of magnitude. Effective utilization and management of data produced at this rate will require totally automated data input, and methods for submission of data by computer network from anywhere in the world. Investigation of hardware and network requirements, design of data structures, database architecture, quality control etc., can be carried out by using a minicomputer, or a workstation network configured to support emulation of the environment that is expected of the fully functioning project.

o Investigate methods for automating the sequencing process, including the possibility of adapting robotic methods that have been developed for analytical and preparative chemistry of radioactive materials.

o Improve techniques for detection and imaging of mixtures of DNA fragments that have been separated on an electrophoretic gels.

Comment: Part of the procedure for sequencing DNA involves identifying the location of molecules on a gel. The current procedure for vizualization is autoradiography. However, techniques exploiting the technology of high energy physics (wire chambers and microchannel plate electron multipliers) are faster and more accurate than autoradiography, and may be ideally suited for incorporation into an automated gel analysis system. National Laboratory expertise in this imaging technology can be rapidly applied to the development of automated genome sequencing methods.

Budget (millions of FY 86 dollars)

1987	1988	1989*	1990	1991	1991
5	10	19	22	22	22

*Assumes purchase of a Cray 2 or equivalent computer with cost prorated over seven years at $6 million/year.