Information on Human Genome Project

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.
PROGRAM TASKS AND RELATED RESOURCES (6/86 - 10/87)

- 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.

- Begin computer simulation 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.

- 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.
- 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.

- 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 visualization 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.

<table>
<thead>
<tr>
<th>Budget (millions of FY 86 dollars)</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
</tr>
</tbody>
</table>

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