Los Alamos
Los Alamos National Laboratory
Los Alamos, New Mexico 87545

DATE April 2, 1986
IN REPLY REFER TO M881
MAIL STOP LS-DO-86-1.2
TELEPHONE 505-667-2690

Dr. Charles DeLisi
Office of Health
and Environmental Research
ER-70, F-208, GTN
Washington, DC 20545

Subject: Genome

Dear Dr. Delisi:

One month has passed since completion of the OHER - sponsored Workshop dedicated to sequencing the entire human gemone, which was held in Santa Fe, New Mexico on March 3 and 4. The Conference was attended by 43 research scientists from the US and Europe of whom 18 were from OHER - sponsored organizations and the remainder from academia and the private sector (please see Appendix I). My aim here is to summarize for you the highlights of this Workshop.

The Workshop was distinguished by a rare and impassioned esprit. Clearly the participants were deeply committed to providing an impartial assessment of the scientific merit and both the socio-political and biomedical impact of so comprehensive an undertaking. The participants were also conscientious in their efforts to provide for you their evaluation comments with respect to the proposed sequencing activity (please see Appendix II).

The workshop participants addressed a series of fundamental questions, which deal with the most compelling issues encountered in the proposed genomic sequencing activity. These were divided into four major segments, which included: 1) sequencing technologies; 2) expected benefits; 3) the architecture and/or coordination of the enterprise; and 4) participants and funding. A strawman agenda (which provided an informal outline) is given in Appendix III. A series of running notes of the actual proceedings were maintained by Dr. David Comings. Although of necessity incomplete, they provide a perspective on the technical proceedings and some flavor and sense of the intense interest generated within and by the discussions (Appendix IV).

I opened the workshop by presenting your charge to the participants to examine the feasibility and potential utility of obtaining the complete sequence of the human genome. Frank Ruddle served as an effective Chairperson for the workshop. His dedicated intensity and broad experience contributed to the focused nature and lively pace of the meeting. He also provided a succinct historical perspective during the Introductory Session.

Technology: Discussions of sequencing technology touched on a broad range of issues including: robotics, the "multiplex" sequencing techniques, stochastic and systematic approaches to the ordering of large DNA fragments, and different technologies for the preparation, separation and cloning of cosmid DNA fragments. There is continuing and rapid improvement in sequencing technology with commercial instrumentation now appearing. The last decade has produced more than an order of magnitude improvement in sequencing speed. Conservative extrapolation of current trends suggests a continuous and accelerating progress in the speed and automation of this technology.

A minor caveat was expressed concerning unforeseen problems in sequencing closure ("the end game"). The possibility is raised that certain sequence domains (e.g. repetitive), might be less amenable to available cloning/-sequencing technologies. The more general perception was that emerging advances in technology would be more than equal to any end game problems. A second caveat predicted that sequencing activities would cluster about areas of basic and clinical interest, resulting in a neglect of "uninteresting" domains. This concern was answered by the perception that inevitable advances in automation and sequencing speed and analysis would readily accomodate such "uninteresting" areas, which are likely to contain important surprises.

Estimates of time and cost clustered about 1.8 to 3.6 billion dollars over a 7 to 12 year period for sequencing the 3.5 billion bases in the haploid human genome. The ongoing and rapid pace of technological advance makes it likely that this estimate is too high. The participants were agreed that cloning and ordering of cosmid or larger fragments should be the first order of business. If not at full scale, sequencing activities should also be encouraged from the start because of the pragmatic value for the future technological evolution of the whole enterprise.

Central coordination and integration of genomic sequencing was seen as essential to minimize duplication and to optimize accuracy and technical excellence. Substantial advantages would accrue from focused investments in sequencing automation, robotics, and supporting computational technologies.

Expected benefits: The participants were virtually unanimous in their strong advocacy for the proposed massive biomedical and biophysical effort, though preceived benefits varied with the perceptions of the beholder. Knowledge of the full organization of the human genome will have a powerful impact on our understanding of embryogenesis, the molecular mechanisms that control gene expression, and especially the molecular bases for inherited diseases. The impact on our understanding of the huge number of neoplastic and malignant disorders will be substantial as will be the benefits for metabolic and cardiovascular diseases. Substantial gains will also accrue for a spectrum of pulmonary fibrosing disorders, the arthridities and especially the auto-immune and immune-deficiency disorders. The complete genomic sequence will illuminate the rapidly growing collection of RFLP's and more perfectly define the extent and clinical import of human genetic heterogeneity. The genomic sequence will also provide very explicit guidance in genetic engineering for human health and for a variety of human health products.

In terms of the sociological and economic effects, the gain in knowledge for human health will markedly improve the quality of human life, expand the duration of an individual's productive years and substantially reduce loss of productivity associated with acute/catastrophic and chronic/debilitating illnesses. The Health Care Financing Administration of the Department of Health and Human Services estimates that the United States public will spend over 400 billion dollars in health care in 1986 (this is exclusive of any investment in biomedical research). The health advantages arising from knowledge of the complete genomic sequence will unequivocally have an impact on these expenditures sufficient to more than offset the costs of the entire enterprise. Even a 1% reduction in health care costs would save more than the estimated cost of the sequencing project. This is above and beyond the program's considerable economic impact, which derives from the creation of biomedical, computational and robotic products spawned by the sequencing effort. Finally, while the participants acknowledge that the information arising from genomic sequencing activity could be misused and misunderstood, they were unanimous in the view that the theoretical and pragmatic benefits would far outweigh the potential for abuse of this information.

Architecture and model of the enterprise:
This topic produced the greatest spectrum of opinion. At one extreme the idea of maximum centralization in a sequencing institute was advocated by Wally Gilbert. Gilbert argued that the huge initial expenditure for bricks and mortar and recruitment would be well and rapidly repaid by the interactive benefits of centralization, critical mass and technical synergy. A blueprint for such an institute is shown in Appendix V. On the other end of the spectrum Sherman Weissman reasoned that a multi-focal effort would provide more generalized research benefits throughout the molecular biology community. Such a multifocal effort would also receive more enthusiatic general support since it would be perceived more as an enhancing rather than a competive enterprise. Most significantly, Weissman reasoned that a multicentric architecture would be more conducive for truly innovative developments, which could provide strong impetus for the activity as a whole.

There was clear unanimity with regard to the need for a very strong and well-conceived central coordinating body, which would perennially and continously identify and evaluate both the most promising approaches and the most appropriate investigators and laboratories to contribute to the entire sequencing effort. There was, throughout the session, strong agreement with regard to the central importance of computational technologies in order to integrate the entire effort and in order to analyze emerging sequences. The call for a computer facility, both to collect, order and distribute the mapping and sequence data, and to analyze the data, met with wide approval. Some estimates concerning the scope and cost of the compuational effort (about 10% of the total cost) are given in Appendix VI in the notes of Christian Burks. There were continual reminders that the enterprise needed to begin in a very flexible mode in view of the current dynamically evolving status both of theory and technology. There were also discussions concerning the merits of sequencing around clinically important foci of interest (e.g., foci

surrounding RFLP's) and concerning the important role to be played by a central coordinating facility with regard to the development of methods for the distribution of cloned DNA fragments both for ordering and sequencing. Looking beyond the human genome, the participants felt that whatever the administrative architecture (more-centralized or multicentric), additional sequencing tasks would emerge relating to a variety of microorganisms and even such mammalian species as the mouse. These extended sequencing activities would continue to provide profound research, biomedical, and clinicial benefits. There was very strong support for the recommendation that OHER proceed to name an Advisory Committee to continue to develop strategies and tactics for the design of the organization, identification of the funding, and development of technologies. It is my belief that virtually without exception those who participated in the Santa Fe workshop would be quite eager to serve you in this capacity.

Funding and Participants:
The members of the Santa Fe Workshop acknowledged readily that the genomic sequencing project would of necessity require multi-institutional and perhaps multi-national support. There are ongoing mapping and sequencing efforts in Europe and Asia that could, when properly integrated, provide a substantial impetus for a worldwide effort in human genome sequencing. The efforts of the Office of Health and Environmental Research are clearly seminal in terms of the DOE's long standing interests in mutagenesis and carcinogenesis, and its ongoing efforts in support of the National Gene Library Project and the GenBank activity. The National Gene Library Project is now entering a second (large insert) phase as a collaboration between the Lawrence Livermore and Los Alamos National Laboratories and this could provide important materials for subsequent ordering, mapping and sequencing activity.

In addition to OHER there is significant interest on the part of the Howard Hughes Medical Institute, which has been interested in funding mapping activities (see Appendix VII). It was also acknowledged that NIH and NSF are cleary interested in these activities as are a number of corporate entities, especially the genetic engineering firms. It was also pointed out that the World Health Organization might constitute one of a number of possible avenues for the development of international cooperation and funding sources. It was emphasized by many of the participants that the funding for this new effort should be clearly distinguished from the funding already available from a variety of health research organizations including OHER and NIH. The necessity for seeking incremental funding was emphasized in the context of making certain that the molecular biology and biomedical research communities would perceive this as a supplementary rather than a competing effort.

In summary, the time is opportune to begin to assemble the funding and personnel who will carry forward the sequencing of the entire human genome. The workshop participants strongly supported the idea that OHER charter a Steering Committee who would assume the initial task of selecting technologies, laboratories, and individuals for initial investment, and also identify those who would assume the long-term stewardship of the entire effort. Your vital support for this extraordinary concept appears at a time when widespread enthusiasm for the activity is beginning to mount in the molecular biology community. It is clear that the participants in this workshop were extremely grateful to you and the Office of Health and Environmental Research for assembling the workshop and expressing strong support for this historic and profoundly important activity. Please let me know how I may be of further service in facilitating this effort.

Very truly yours,

Mark W. Bitensky M.D.

Mark W. Bitensky, M.D.
Senior Fellow
Los Alamos National Laboratory