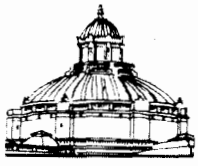
Issue Brief

Order Code IB88012

PROPOSAL TO MAP AND SEQUENCE THE HUMAN GENOME

Updated February 2, 1988



by

Irene Stith-Coleman

Science Policy Research Division

Congressional Research Service

CONTENTS

THE PROPOSAL TO MAP AND SEQUENCE THE HUMAN GENOME

SUMMARY

Recent advances in molecular biology have made it possible for scientists to discover the genetic makeup of humans. This capability has led to the suggestion of some experts that the entire human genome be sequenced and mapped. This would involve determining the linear order of the 3 billion units that make up the human genome, called nucleotides, and locating the functional parts of the genome known as genes. In 1986, the Department of Energy (DOE) proposed that this be done in a large-scale centrally coordinated project. If approved, this would be the largest biological science project ever undertaken, and would cost an estimated few hundred million to a several billion dollars.

The information obtained from mapping and sequencing the human genome could be of substantial benefit for treating the 4,000 genetic diseases known to afflict humans. In addition, it could have enormous potential for commercial applications in fields such as pharmaceuticals and agriculture. Indeed, some supporters view the proposed initiative as being of significant value in helping the United States maintain Its competitive edge in the area of biotechnology. However, the proposed project has led to a great deal of debate among scientists. Central to this discussion is the belief that significant improvements are needed in the technologies required to map and sequence the entire genome.

Another issue raised is how the genome should be mapped and sequenced. Some experts fear that initiation of a major centralized project to sequence the entire human genome could result in funds being diverted from other crucial research projects. Such scientist believe that a first priority should be given to mapping and sequencing only the portions of the genome that have known biological significance (encoded by only about 5% of the genome's DNA). Contemplated in this option is an incremental selective series of projects with reduced costs spread over a longer period. Advocates of a major genome initiative argue that the so-called "nonfunctional" DNA in the genome may have important unidentified functions, which could be discovered by sequencing the entire genome.

A number of public policy issues will be presented by future research on the human genome, whether done in a large-scale project or a series of smaller projects. These include: ownership of information questions, such as copyright and patent rights to genetic information; ethical and legal concerns related to how the information will be used; and implications to international competitive issues, including the role of foreign countries in the production and use of human genome information.

ISSUE DEFINITION

Scientists are now able to discover the entire genetic makeup of humans. This can be done by determining the linear order of the 3 billion units that make up the human genome, called sequencing, and identifying the location of the genome's functional parts, or genes, known as mapping. The Department of Energy (DOE) has proposed that the entire genome be mapped and sequenced in a large-scale centrally coordinated project. If approved, this would be the largest biological science project ever undertaken. It would cost an estimated few hundred million to several billion dollars, and has potential for producing major scientific, medical, and commercial benefits. Initial issues for Congress are: Should Federal sponsorship of such a "big science" project be authorized and funded? If so, under what terms, controls, and agency leadership? Or, should genetic research proceed in the "small science," less comprehensive mode that has produced the progress of recent decades? Applications of human genome information obtained be either approach will present a number of bioethical, commercial, property, and legal policy issues.

BACKGROUND AND ANALYSIS

Scientific Background

Research in genetics, the science of heredity, historically has been driven by efforts to treat genetic diseases in humans. Genetic diseases are believed to be the most widespread of all disorders, accounting for some 4,000 known human disorders, many of which are lethal. The quest for remedies has led to significant advances in genetics, particularly during the past decade, and has given scientists the capability of discovering the entire genetic makeup of humans. Such information could be very beneficial in the search for the causes of genetic diseases. This prospect is especially significant because in most cases, no cure is yet available for genetic diseases and doctors can offer only palliative treatments, if any.

The key to unraveling the mysteries behind genetic diseases appears to reside in the genes of the human genome. The genome, which is found in each cell, contains the genetic code that directs the entire makeup of humans. The human genome consists of 23 pairs of chromosomes, each of which is made up of a double-stranded DNA (deoxyribonucleic acid) molecule containing long strings of four chemical groups called nucleotides, or bases. These nucleotides, totalling an estimated 3 billion pairs, are arranged in various largely unknown combinations. Dispersed among them are segments of nucleotides that form the genome's estimated 100,000 functional units, known as genes. However, these estimated 100,000 genes are encoded by only 5% of the 3 billion nucleotide pairs that make up the genome; the remaining pairs have no known function.

Genes are hereditary units that are transferred from parents to child in a generally predictable fashion. Activities originating from these

hereditary units govern numerous processes including the growth, structure, function, embryonic development, and possibly even the aging process of an organism. For example, proteins, whose production is orchestrated by genes contained in the genome, make up the structural components of bone, muscle, brain, blood, and other human tissues, and regulate most activities of all living organisms.

Genes are believed to range in size from less than 1,000 nucleotides to as many as 1,000,000. Many structural and functional abnormalities in animals and plants are known to be the result of subtle changes in the arrangement of the nucleotides making up a gene. When some external factor (such as radiation) or some internal factor (such as an accidental deletion or a shift in the position of a nucleotide) occurs, the result may be modification of the genetic code and a change in the functioning of the organism. Severe effects in humans can result from abnormalities in genes, as seen in individuals with conditions like sickle cell anemia, hemophilia, and Huntington's chorea. Moreover, recent discoveries indicate that defective genes may play an important role in more prevalent diseases such as heart disease, arthritis, cancer, and Alzheimer's disease, which previously were not thought of as being hereditary.

Motivation for Mapping and Sequencing the Genome

In the past, the approach taken to mapping and sequencing the human genome was that of identifying and determining the order of the nucleotides in genes of biological interest, typically those related to specific genetic diseases or specific human functions. For example, research has been aimed at locating the genes responsible for diseases like sickle cell anemia and Huntington's, and the genes involved in protecting humans from infections. In this approach, scientists have used recently developed techniques like recombinant DNA to identify, or map, more than 1,300 of the estimated 100,000 genes contained in the human genome. To date, about 500 of these genes have been partially sequenced, but only 12 human genes have been completely sequenced. The approximate location of at least 2,000 other genes also has been identified with modern biological tools. Doctors already are using these findings to help diagnose about 30 human genetic diseases during the first 3 months of pregnancy. However, additional research is necessary before they will be able to treat any of these genetic diseases.

In addition to the human health aspect, congressional interest in a major initiative to sequence and map the human genome has been heightened by the potential commercial applications of the information obtained through biotechnologies. The field of biotechnology is now in an infant state; however, the industry is expected to grow dramatically over the next several years when genetic developments will significantly increase the number of products, particularly pharmaceutical and agricultural. Genetic research already has provided numerous biological tools, including recombinant DNA, that have enormous commercial potential. Commercial application of such techniques are evident with the production of pharmaceutical and animal drugs like insulin, human growth hormone, and bovine growth hormone.

Currently, the United States is in a good position to capitalize on the biotechnology industry because it is the generally accepted leader in most of its applications. It is unclear, however, whether the specialized information and "spin-off" technologies likely to be obtained from mapping and sequencing the human genome will be of sufficient economic value to justify initiating a "big science" project, a central issue related to the proposed large scale genome initiative. An important consideration is that the commercial payoffs from mapping and sequencing the human genome are not expected to occur until many years after a major genome project has been initiated.

Another issue that has been raised concerning possible Federal support for a major genome mapping and sequencing project is the role of financial support from private industry, which probably will reap many of the commercial benefits. The private sector, to date, has expressed little interest in funding a major human genome mapping and sequencing initiative. Many of the research efforts presently being supported by private industry are targeted at specific portions of the genome that are related to known human diseases or functions. This approach is viewed as having the greatest near-term commercial potential. Many U.S. biotechnology companies are relatively small and cannot afford to carry the economic burden required to map and sequence the entire human genome. This factor is particularly significant because major human genome mapping and sequencing efforts are currently being funded abroad in countries like Japan. Much of this support comes from large corporations who can better afford the required long-term investments. Factors such as these, as well as the prospect of the biotechnology industry becoming a multibillion dollar market in the near future, has led to concern that the United States could lose its lead in areas like the pharmaceutical and agricultural industries to international competitors like Japan if a major American-sponsored genome project is not initiated.

Mapping and Sequencing The Entire Genome

One approach to mapping and sequencing the entire human genome involves creating what's known as a physical map of the genome. To produce this map, the entire genome will need to be broken into overlapping fragments, which would have to be cloned (multiple copies made) and arranged in the linear order in which they appear in the intact genome. A complete physical map will show the exact distance in nucleotide pairs between points, or landmarks, as they appear on the genome. Once the fragments have been correctly lined up, the order of the nucleotides within each fragment can be determined, or sequenced. This information will have limited importance, however, without another map, known as a genetic or genetic linkage map. To create this map, samples of genetic material (extracted from blood samples) must be taken from individuals of large families representing several generations. A complete genetic map will show the correct order and identification of the estimated 100,000 genes as they appear on the 23 pairs of chromosomes of the human genome. As mentioned previously, these 100,000 genes are encoded by only 5% of the nucleotides contained in the entire human genome.

Numerous resources would be needed to map and sequence the entire human genome. These include techniques to map and sequence genetic material, as well as computer hardware and software to store and analyze the massive quantity of genetic material. While some of these resources are available, improvements are needed and additional technologies will have to be developed. Since the Federal Government, a major supporter of genetics research in the past, is the principal owner of many of the existing resources, Federal support may be the most effective way for the United States to devote the necessary resources to a concerted genome project. However, in an era of serious Federal resource constraints, many are concerned about the possible deprivation of resources in other areas of research (agricultural biotechnology, other biological research) if a "big science" genome project is federally funded.

Controversy Over A Major Genome Initiative

There has been no fundamental disagreement about the importance of knowing the location and the precise chemical structure of genes, which could be obtained from a major genome initiative. Much of the debate among biologists, however, has been focused on the merits of initiating a major centralized effort to sequence the entire genome. To date, the majority of the information obtained on human genes was gathered by scientists studying specific diseases, often in a relatively small scale peer-reviewed research environment. The National Institutes of Health (NIH), the major supporter of biological research in the world, has funded most of this research. A large-scale centralized genome project would not only change this research approach significantly, but could also alter NIH's status in this area. Many researchers fear that the massive resource requirements of a major genome project would cause adverse effects on other important biological research, and that research funds would be diverted from other crucial research programs.

Although technology currently exists to map and sequence the human genome, much of it is in an infant state of development. Many scientists, therefore, believe that additional technology development is needed before a major centralized project should be initiated. If a major genome mapping and sequencing project was initiated with existing techniques, scientific experts estimate that it not only would cost several hundred million to several billion dollars, but also would require the participation of hundreds of scientists. In addition, many of these researchers claim that the resulting data would have a high error rate. These experts believe that the resource requirements could be significantly reduced if the first research priority is given to the automation of mapping and sequencing technologies, improving data storage and analysis methods, and increasing the support levels for current gene mapping research projects. If this approach was taken, they argue that the total cost could be reduced by a thousand-fold and the error rate could be decreased by at least ten-fold assuming that a major effort was directed at these areas over the next several years. (Leroy Hood and Lloyd Smith, Genome Sequencing: How to Proceed, Issues in Science and Technology, Spring 1987: 36-46; and Roger S. Johnson, Ph.D., The Human Genome Project: What Impact on Basic Research?, FASEB Office of Public Affairs, December 1987: 502-505).

Some opposition to a major genome sequencing project is based on the fact that 95% of the 3-billion-unit genetic code has no known biological function. These critics argue that efforts should be focused instead on locating (mapping) and studying only the functional genes and their products (proteins), and that this should be done in the diverse manner that has been used in the past, rather than as a major centralized project. This research approach would be directed at studying genes associated with known diseases or human functions.

Proponents of a major centralized genome project counter that the so-called nonfunctional DNA contained in the human genome perhaps have important functions not yet identified which may be resolved by sequencing the entire genome. Moreover, advocates for this project argue that better management and organization of available technology may be a greater priority for sequencing and mapping the human genome than developing additional technology. This approach would significantly improve the time and money requirements, according to Dr. Walter Gilbert of Harvard University, a leading proponent of a major genome initiative.

Current Activities on Genome Research

Federal Agencies

The Federal Government currently is the major supporter of genome related research in the United States. Much of this work is administered by the National Institutes of Health (NIH) and the Department of Energy (DOE). NIH, the principal supporter of biological research, is spending about $300 million each year on gene sequencing and mapping related research. Approximately $90 million of these funds are focused on human genome research and support more than 3,000 peer-reviewed research projects. Congress has appropriated an additional $17.2 million to NIH for FY88 to expand research efforts in gene mapping.

NIH also administers and partially funds (with DOE and NSF) GenBank, one of the major DNA sequence databases that are available. GenBank is a international computerized database that records, stores, and disseminates DNA sequence data. It is located at Los Alamos National Laboratory and is closely linked to the other major DNA sequence database, the European Molecular Biology Laboratory in Heidelberg, West Germany.

The National Library of Medicine has been involved in developing databases that contain DNA data important for mapping and sequencing the genome. One of these databases, On-line Mendelian Inheritance in Man (OMIM), is a computerized version of the book which has the same name that was created by Dr. Victor McKusick of John Hopkins University in Baltimore, MD, a pioneer in the area of genetics. OMIM is a gene mapping database that contains disease related DNA information. This database is also linked to the Human Gene Mapping Library owned by the Howard Hughes Medical Institute. NLM also administers the recently established National Biotechnology Information Center. This center will serve as a repository for storing and disseminating genetic information, and also as a place for

coordinating and supporting existing genetic databases. The National Biotechnology Information center was appropriated $3.83 million for FY88.

DOE, who first formally proposed a major centralized human genome project, plans to spend $12 million during FY88 on human genome-related research. Most of this work will be targeted at applied biology and engineering necessary to do sequencing and mapping. In the past, much of the genetics research supported by DOE was aimed at examining the mutagenic effect of environmental contaminants like radiation on human genetic material. As a result, DOE employs scientists who have expertise in disciplines like molecular biology, engineering and mathematics. DOE's expertise and resources are reflected in the agency's proposal for a primary genome program for FY88, which will be aimed at developing instrumentation, mathematical, and other aids for sequencing and mapping. As mentioned earlier, DOE also partially funds the GenBank database.

Other Federal agencies also are involved in supporting research in genetics, or are funding programs in areas that would support a genome mapping and sequencing effort. The National Science Foundation (NSF) funds research to develop new instrumentation for molecular biology. It also partially supports the GenBank database. The U.S. Department of Agriculture (USDA) has an increasing commitment to genetics research, but their research efforts are focused primarily on organisms of interest in agriculture, i.e., affecting animals and plants. However, much of this work, particularly that on animal diseases, has a very high relevance to similar work on human disease. In addition, other agencies like the National Bureau of Standards and the Centers For Disease Control are already involved in functions like quality control that will be important to the genome project. Coordination and collaboration among agencies are important, and this has been initiated between several agencies, including NIH, DOE, and NSF.

Private Sector

Two companies are largely responsible for a major portion of the human genome related research currently being supported by the private sector. They are Collaborative Research Inc., a biotechnology company located in Bedford, MA, and the Howard Hughes Medical Institute (HHMI), a non-profit organization (executive offices located in Bethesda, MD), the world's largest source of private funds for health and biomedical research. Many of the efforts supported by both these organizations are targeted at mapping genes related to specific diseases. Collaborative Research recently announced the completion of its "primary genetic linkage map" and HHMI scientists are expected to complete a similar map soon.

These genetic maps make use of segments of DNA called restriction fragment length polymorphisms (RFLPs) that are used as markers in the human genome. The genome of each individual, except for identical twins, differs from another person's genome by perhaps one-tenth of a percent of their nucleotides. If chromosomes from two people are cut using special enzymes that will snip the DNA at known points, some of the pieces of DNA fragments produced will differ in lengths; these variations are known as RFLPs. Certain RFLPs show up only in the DNA of people with an inherited disease like Huntington's and cystic fibrosis, and are located near the

defective gene. These characteristics make RFLPs very useful markers to help diagnose associated genetic diseases, as well as to help pinpoint the location of associated genes.

The Collaborative Research genetic linkage map consists of 404 markers located along the human genome. The company has already announced its intent to seek patents for any of their identified markers that are used to diagnose a disease, as well as for the site on the genome of any gene that the company's scientists identify using its primary genetic map. This approach is viewed by the company as being necessary in order to protect its proprietary rights while also making information available to other scientists. However, researchers have raised the concern that the company has not made its data available except in a few provisional cases. Scientists believe that they should have quicker access to Collaborative's Research data, which could accelerate the process of searching for treatments for genetic diseases.

Howard Hughes Medical Institute-supported scientists have a strong collaborative relationship with biologists working in the area of genetics. This Institute (HHMI) also owns important human genome related resources; for example, HHMI owns the Human Gene Mapping Library (HGML), the largest existing computerized database for the human gene map. This database is linked to the OMIM database, as mentioned above, and is accessible to researchers across the world at no charge. Scientists can use this computerized system to obtain available information on DNA associated with hereditary diseases, including location in the genome of any gene or marker that has been identified, who did the work, and any related references.

HHMI scientists are also developing a genetic linkage map. This map, which is anticipated to be completed soon, reportedly contains 475 markers. Genetic linkage maps will significantly speed up the process of identifying the 100,000 human genes. However, the markers located so far are separated by 10-20 million bases, or more. Thus, the job of zeroing in on human genes remains difficult even with the information contained in existing genetic linkage maps. As research in this area continues to progress, genetic linkage maps are expected to become more detailed, containing several thousand markers, only about 1 million bases apart, over the entire human genome. Once the resolution of these maps improves, scientists should be able to locate the genes themselves more quickly, which is required in order to make a complete genetic map. Researchers supported by both HHMI and Collaborative Research, as well as those being funded by the Federal Government are actively attempting to create more detailed genetic linkage maps.

Presently, sequencing of the human genome is less developed than mapping, with less than 1% of the human genome having been sequenced to date. While many scientists believe that the genome should not be sequenced until technologies are improved, at least one scientist, Nobel Prize laureate Walter Gilbert, advocates initiating genome sequencing now with the technologies already available. Dr. Gilbert recently announced his intention to sequence the human genome as a private venture by forming a biotechnology company called Genome Corporation. Although Gilbert has not yet raised all the required funds needed to start his

company, he has already stated his plans to copyright the genome sequences that he determines, and sell this information to researchers who wish to use it. These plans have evoked debate among many scientists who believe that genome information should be published and remain in the public domain.

International Activities

Current congressional concern with trade imbalances has focused attention on Japanese activities and policies in biotechnology. Several of the technologies needed for mapping and sequencing are under development in Japan. For example, research targeted at developing automated sequencing devices was underway in 1986, according to Dr. Akiyoshi Wada, a pioneer proponent of the Japanese human genome mapping and sequencing efforts. (A sequencer developed in the United States is used in DOE laboratory work.) In Japan, the Science and Technology Council is supporting studies on a super-DNA sequencer project, started about 1981. This project involves the development of robotic techniques. The project has incorporated existing technologies as well as developing new techniques. For example, the Fuji Film Company had developed a special film to meet the requirements for electrophoretic separation processes. These films are now being sold commercially in the United States and will be incorporated in the process described by Wada. Also, methods developed in the United States and England for preparing DNA for sequencing are being used in the Japanese devices assembled by Hitachi Software Company and by Seiko. Wada indicates that DNA sequencing should be accomplished in a sequencing factory equipped with automated systems and that the sequencing should be coordinated at an international center open to all researchers. He is said to favor having the center located in Japan. Wada views this work on an automated sequencer as an opportunity for Japanese companies to use their strengths in electronics, computers, and materials sciences in the field of biology. Seiko, for example, is said to have built a new life sciences instrument division within the corporation. Wada hopes to have his sequencing project finished in 1988 and operating in 1990.

While the work in England is not on the same financial scale as the United States research program, the work is of high quality. As an example, a major contribution is being made by an English group that has almost completed the physical mapping of the entire genome of a roundworm. As another example, the Imperial Research Cancer Fund in London recently identified research that has located the site of a gene deletion related to colon cancer.

Biomedical research in a number of French laboratories also involves activities on gene mapping and sequencing. In terms of international participation, probably one of the best known efforts is that of the Human Polymorphism Study Center in Paris. This French-based international organization serves as a reference depository of important genetic material that is accessible to researchers across the globe. A unique arrangement of this organization, which requires long-term commitment and cooperation, has been resolution of the problem of data sharing. The Center requires participants to share the data they secure from the cell

lines provided but the participating laboratories are not required to reveal data on the specific processes which are used.

In West Germany, a primary facility of importance in the international exchange program is the European Molecular Biology Laboratory (EMBL) at Heidelberg. EMBL participates (via its Nucleic Sequence Data Library) in the GenBank project. About 50% of the data in this project is said to be coming from Europe. Automated sequencing devices are also under development.

Data Management and Analysis

Despite professional disagreements about the DOE total mapping and sequencing proposal, there is general consensus that improvements are required in managing the information systems already accumulating data from scattered mapping and sequencing activities. Such data include collections of: DNA in "libraries" and cell "banks"; biological "probes" used also for diagnostic and genetic studies; multiple family cell lines used in collaborative genetic studies; biological "vectors" used for recombinant DNA; and similar materials resources.

The sequencing data are expected to increase at a rapidly accelerating rate in the near future. While the quantity of sequencing data is large, it is not unusually so in the information processing field. The task, however, of processing sequencing data to facilitate analysis is more complicated than most other areas, demands special software programs for analysis as well as high-speed large capacity computers with networking capabilities.

Some researchers suggest establishing a new term, "biotechnics," for the use of computer technologies for all biological disciplines. Processing of the DNA code has been described as being very similar to the processing and analysis of words from commonly used languages. Thus the devices which sequence nucleic acids or proteins are "reading" a language except that the language -- in the case of the genome -- evolves from only four letters symbolizing the four bases present in the sequence. Since programs are available for reading and storing languages, and in the case of cryptoanalysis, logic for coding and decoding, it is conceptually possible for computer programmers to design the programs necessary for storing biological data in a form permitting retrieval of specific sequences and various manipulations and comparisons. This will not be an easy task, however, because of the need to retain access to data filed with different programs in a number of databases. At the current stage of computer utilization, geneticists are still working at the early binary level of computer languages. Geneticists will soon need a more effective computer assembly language and will eventually require a genetic programming computer language.

Existing information systems rapidly are becoming outmoded for current needs in genetics. The National Library of Medicine's large computerized database is of immediate relevance to all mapping and sequencing projects. The data management system which has evolved at NLM to handle reference and other published documents is accessible all over

the world. Dr. Donald Linberg of NLM suggests that there is a need for new systems of data entry, standardization of nomenclature, and an urgent requirement for more effective systems of data retrieval and analysis.

The Howard Hughes Medical Institute (HHMI) has initiated efforts to improve the interface between a variety of data and materials bases providing information to investigators in genetics. In fact, it is possible that HHMI may become a major private facilitator of the coordination of many private DNA databases. HHMI's interest in this task has precedence in The Gene Mapping Library supported by HHMI at Yale.

It may not be logical to embark on a major sequencing project to produce large quantities of essential data until the information system into which such data would be entered is adequate for the task. GenBank, the database that shares information with the European Molecular Biology Laboratory, appears to be unable to keep up with data entry tasks as the amount of information exceeds available manpower. If researchers are to have a successful interface between the descriptive databases associated with collections of DNA, probes, and other materials and the databases filing research information, the information must be stored and retrievable quickly and accurately. Many of the interfacing program languages now used are quite cumbersome. Journal editors can no longer cope with the mass of sequencing data and, since these data are not being printed, there must be access to other resources for investigators to use for filing the results of their work. Such systems must not only be readily accessible but open to all collaborators. This implies an expanded level of international coordination.

Although there are some databases funded primarily by foreign or private funds, most of the systems of major significance are supported by Federal funds. For this reason, it may be possible to secure the desired levels of coordination and standardization by an increased effort in Federal programs. All of the agencies associated with databases relevant to programs in molecular genetics would be involved. The Bureau of Standards also could be included, since this agency is already working in computer language standardization activities.

Other Policy Issues

Work in the field of biotechnology has raised a number of ethical and legal issues. One major concern about mapping and sequencing the human genome involves the potential uses of the information obtained from such as effort. While the actual process would not create significant ethical or legal issues, the availability of a physical map or gene sequences could bring processes that would raise these issues much closer.

Few would argue that an important benefit of such information would be the ability to better understand and someday treat genetic diseases. Knowledge gained through mapping and sequencing efforts, however, will make it possible to predict individuals' susceptibility to diseases, many of which would not occur until later in life. The potential use of this information, so-called "predictive-genetics," raises significant ethical, social, and legal concerns. While an individual may want to know his/her susceptibility to an illness such as heart disease, he may not wish to know that he carries a defective gene that will cause a nontreatable fatal

disorder like Huntington's disease. In the first case, the risk of heart disease may be reduced by environmental factors such as dietary modification, which is within the individual's control. However, in the case of a disease like Huntington's, the individual who carries the defective gene currently can do nothing to prevent the disease and it cannot be effectively treated by doctors. Of still greater concern is the potential use of predictive genetics by health and life insurance companies, and employers, to withhold insurance or charge exorbitant insurance rates based on individual disease susceptibility.

Another major ethical concern raised by the availability of mapping and sequencing data is the possible use of such information to screen the genetic makeup of unborn children. Some predict that the discovery of genetic defects will lead to an increase in the number of abortions performed. Data on currently available genetic tests for fetuses indicate that discovery of defects leads to an abortion in many cases. Although many genetic defects can be identified in fetuses, no therapeutic treatment is now available to correct most of these defects. This situation is not expected to improve in the near future.

A major legal issue surrounding the human genome mapping and sequencing proposal is whether the information obtained should be owned by private entities through patents, copyright, or other intellectual property law. A significant amount of the mapping and sequencing data is expected to be commercially valuable because it will aid in the development of products such as therapeutic drugs. Some private companies are presently involved in mapping and sequencing the human genome with a major objective being commercial application of the information obtained to detect and treat specific diseases. What role should private companies play in mapping and sequencing the human genome, and what rewards, if any, should these companies receive as incentives to participate in these efforts are unresolved issues. Representatives of the private sector already have voiced their intentions to copyright, patent, and sell genetic data generated from mapping and sequencing the human genome.

Opponents of private ownership of genetic information believe that it will impede access to and distribution of scientific information that in the past has been published for general use. Moreover, some believe private ownership will make scientists involved in genome mapping reluctant to share data. For example, concern has been raised that important steps in mapping and sequencing genetic material for a disease such as cystic fibrosis could be held confidential until all the steps are resolved. It is possible, however, that if such preliminary information was made available to other scientists, the entire project could be completed faster. A major consequence of this practice could be a delay in the detection and treatment of human diseases.

OTA and NAS Studies for Congress

The Office of Technology Assessment is currently conducting a study for the Congress on mapping and sequencing the human genome. Several issues will be highlighted: the controversy about the proposal for a major centralized approach; research funding requirements; the need for more coordination; legal issues concerning patents and copyrights;

technology transfer; international cooperation; and economic competitiveness. It is anticipated that the report will be released about March 1988.

The Board on Basic Biology, Commission on Life Sciences, National Research Council, National Academy of Sciences, has recently published a report entitled "Mapping and Sequencing the Human Genome." The National Academy of Sciences (NAS) committee recommended, among other things, that a major human genome mapping and sequencing project be initiated, and that the greatest priority be given to developing new mapping techniques. The committee also recommended that $3 billion be earmarked for this genome initiative over a period of 15 years at a rate of $200 million annually, and that the research should be done by individual scientists and "medium-sized" research teams, as opposed to a large centralized project. The committee based its recommendations on the potential medical applications of the data and not on the issue of its potential impact on the competitiveness of the U.S. biotechnology industry.

Policy Options For Congressional Consideration

Congress has already taken steps to expand support for gene mapping and sequencing efforts. Examples of this include the creation of the National Biotechnology Information Center at NIH, and increased funds for NIH-supported gene mapping research. In addition, DOE has undertaken an initiative aimed at mapping and sequencing the human genome. Some observers believe that the current level of genome support is adequate. However, should Congress determine: to support further expansion of genome related programs, to promote the competitiveness of the U.S. biotechnology industries, and/or to address genetic diseases, it may consider the following options:

(1) Keep the current human genome research program, but increase the funding available for peer-reviewed grants. The existing research program, for the most part, is based on research proposals approved according to merit. The major objective of this program, as in private sector programs (e.g., Collaborative Research), is directed at mapping and sequencing pieces of the human genome that have known biological significance. An example of this approach would be funding a project aimed at mapping and sequencing the gene for Huntington's disease. The primary commercial application of this kind of information would likely be in the pharmaceutical industry.

(2) Increase resources to existing Federal research programs relevant to genome mapping and sequencing in all organisms of economic importance to the United States, including agricultural plants and animals (livestock, plant commodities, agricultural pests). The commercial applications of this kind of information would be broad-based, and would probably include human health and agricultural products.

(3) Expand existing research priorities to include initiatives directed toward technologies that will improve the efficiency of genome mapping and sequencing. These initiatives would also include improving data integration and coordination so that genome information could be better utilized once collected. This option might be coupled with options (1) and (2) above.

(Note: The research priorities listed in (1)-(3) above might be accompanied by the establishment of a coordinating body to facilitate communication between representatives of different research programs and to harmonize data collection, as well as monitor overall progress.)

(4) An alternative to the approaches discussed in the above options would be to initiate a new centralized effort aimed at sequencing and mapping the entire human genome. Program administration, including establishment of research priorities, would occur through a principal central body. DOE, which advocates this type of program, proposes the development of hardware and software to improve the efficiency of genome mapping and comprehensive sequencing of all the 3 billion nucleotide pairs in the human genome rather than a focus on only those segments of known biological significance.

LEGISLATION

H.R. 393 (Pepper)

Establishes the National Center for Biotechnology Information within the Department of Health and Human Services as a component of the National Library of Medicine. Introduced Jan. 6, 1987; referred to Committee on Energy and Commerce.

H.R. 3006 (Lujan)

Establishes the Council for Research on Enabling Technologies, the National Policy Board on the Human Genome and the Human Genome Consortium, and a regional Institutes for Entrepreneurial Studies. Introduced July 23, 1987; referred to Committees on Armed Services, on Energy and Commerce, and on Science, Space, and Technology. Referred to Subcommittee on Energy and Power Aug 3, 1987.

S. 1480 (Domenici)

Establishes the Council for Research on Enabling Technologies and the National Policy Board on the Human Genome and the Human Genome Consortium. Introduced July 10, 1987; referred to Committee on Energy and Natural Resources. Referred to Subcommittee on Research and Development Sept. 17, 1987.

S. 1966 (Chiles)

Amends the Public Health Service Act to improve information and research on biotechnology and the human genome, and other purposes. Introduced Dec. 18, 1987; referred to Committee on Labor and Human Resources. Executive comment requested from Departments of Health and Human Services, Agriculture, Defense, Commerce, and Energy, and OMB Jan. 15, 1988.

CONGRESSIONAL HEARINGS, REPORTS, AND DOCUMENTS

U.S. Congress. House. Committee on Appropriations. Energy and water development appropriations bill, 1988. June 17, 1987. Washington, U.S. Govt. Print. Off., 1987. p. 76-69. (100th Congress, 1st session. House. Report no. 100-162)