chapter 1

Summary

CONTENTS

Figure

Table

Chapter 1

Summary

"We want the maximum good per person; but what is good? To one person it is wilderness, to another it is ski lodges for thousands. To one it is estuaries to nourish ducks for hunters to shoot at; to another it is factory land. Comparing one good with another is, we usually say, impossible because goods are incommensurable. Incommensurable cannot be compared.

Theoretically this may be true; but in real life, *inconnnensurables are commensurable*. All that is needed is a criterion of judgment and a system of weighing."

Garret Hardin, "The Tragedy of the Commons, " *Science* 162:1243-1248, 1968.

"Congress is the place where we make impossible choices between apples and oranges. We do it every year in preparing the largest budget on the planet."

Congressional staff member, 1988.

"All legislative powers granted shall be vested in a Congress of the United States . . . . No money shall be drawn from the Treasury, but in consequence of appropriations made by law. . . ."

Article 1, U.S. Constitution.

The mysteries of inheritance are surrendering to modern biology. Over a century ago, Austrian monk Gregor Mendel demonstrated that the inheritance of traits could be most simply explained if it were controlled by factors passed from one generation to the next. These units of inheritance came to be called genes. The complete set of genes from an organism is called its genome. Some traits are best explained by inheritance of single genes (e.g., many genetic diseases, colorblindness), but most, including many nongenetic diseases, involve combinations of multiple genes with environmental factors.

Scientists discovered in the 1940s that genes consisted of DNA (deoxyribonucleic acid), and in the 1950s they further elucidated the mechanisms of inheritance. In 1953, Watson and Crick described the structure of DNA—the double helix—which provides at once an explanation of how genetic material is inherited and how genes direct cellular function. DNA encodes the blueprint for every living thing; it is packed into chromosomes which can be seen under a light microscope. The genome of an organism can thus be defined as the DNA comprising its chromosomes. Each human cell has 46 chromosomes in 23 pairs. One chromosome of each pair is inherited from each parent. DNA consists of long chains of chemicals called nucleotide bases. There are four such bases, represented most simply as A, C, T, and G. The order of bases making up DNA is called its sequence. The DNA sequence contains the instructions that specify the production of molecules, usually proteins, that provide cellular structure and perform biochemical functions in the cell.

Our understanding of genetics has advanced remarkably in the last three decades as new methods of manipulating and analyzing DNA have been developed. Recombinant DNA technology enables scientists to insert DNA from one organism directly into that of another, thereby allowing them to study how genes function in relatively controlled conditions. New methods to detect and purify small amounts of DNA, new techniques to handle and analyze DNA that is millions of bases long, and novel scientific instruments have augmented the tools scientists use to understand heredity. These powerful and rapidly evolving technologies have provoked debate in recent years about whether and how to mount a concerted research program to map the human genome and to determine its DNA sequence.

To date, the combined efforts of government agencies, university researchers, and private sup-

porters of biomedical research have produced rough but extremely useful maps of DNA markers covering most regions of the human chromosomes. Chromosomal locations of over 1,215 human genes are now known (of the 50,000 to 150,000 estimated to exist), including those causing all 20 of the most common genetic diseases. Sequencing of DNA from human beings has increased sharply in recent years, yet far fewer than 1 percent of the more than 3 billion bases comprising the human genome have been sequenced (see figure l-l). The function of only a few hundred human genes is known. Some genetic disorders are understood at the molecular level (e.g., sickle cell disease and Tay-Sachs disease), but the mechanisms underlying most genetic diseases remain unknown. Genetic factors underlying other diseases are known only in barest outline.

The growing power and speed of research in molecular biology have led to proposals to apply novel molecular biological methods to the genetics of entire organisms. **Research and technology efforts aimed at mapping and sequencing large portions or entire genomes are called genome projects.** These proposals would build on experience already gained from mapping lower organisms (e.g., yeast, nematodes, and bacteria) and sequencing some virus genomes and regions of other organisms, yet they would be more ambitious in scale and complexity. More specifically, a public debate began in 1985 about the feasibility of mapping, and perhaps sequencing, the human genome and that of certain other organisms. The debate has often been cast as an on-off decision about whether there should be a concerted Federal effort, yet this is an oversimplification. There are many component projects at different stages of completion: Systematically making maps of human chromosomes is a continuation of ongoing efforts, for example. Databases for genetic information and repositories for research materials are essential whether or not there are other special efforts. Developing new technologies is widely agreed to be important and will require focused research programs. The most contentious issue is whether the DNA sequence of all human chromosomes should be determined. There is little doubt that large regions of human chromosomes will be sequenced eventually, but there is vigorous debate about whether a massive, concerted sequencing effort is warranted. This remains an open question that is likely to be resolved only after pilot projects to determine the sequence of other organisms, small human chromosomes, or chromosomal regions of special interest have been performed. Pilot projects can demonstrate the technologies and should also determine whether dedicated sequencing efforts are efficient and scientifically sensible.

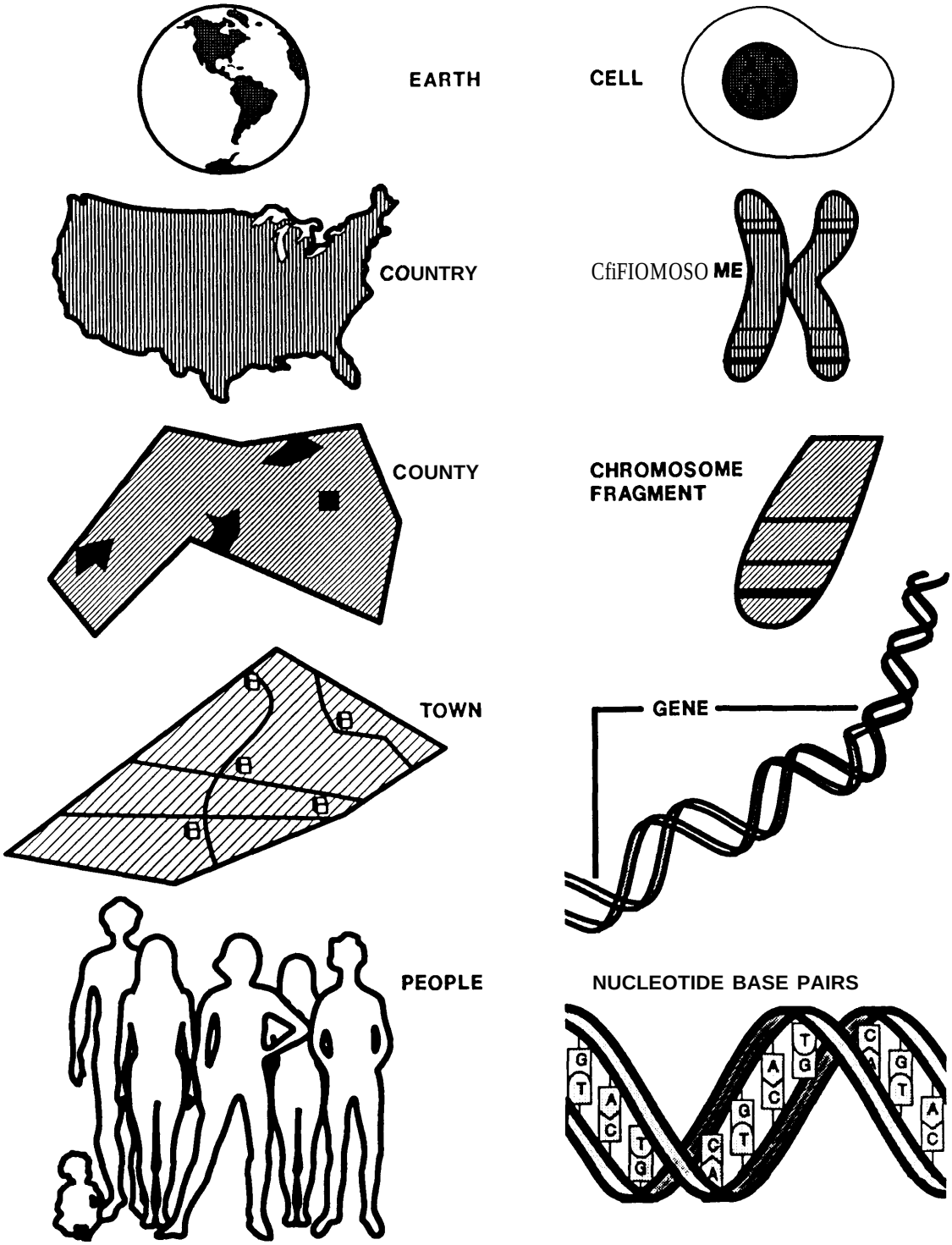
Two scientific advisory groups-one reporting to the Department of Energy (DOE) and the other convened by the National Research Council (NRC) of the National Academy of Sciences—recommended augmented funding of $200 million per year for genome projects. An Office of Technology Assessment (OTA) workshop attempted to estimate the costs of major component projects. Projections fell into the range of $45 to $50 million per year initially, increasing to $200 to $250 million per year over 5 years. Funding recommendations made by the scientific advisory committees would cover most but not all costs estimated by OTA.

DEBATES ABOUT MAPPING THE HUMAN GENOME

The debate about mapping the human genome can be traced through several phases. Until the 1960s, techniques for locating human genes were rudimentary, and human genetics was based pri - marily on analysis of inheritance patterns of diseases and other observable traits through family trees. In the late 1960s and through the 1970s, scientists developed the first maps of human genes, based on direct observation of chromosomes. In successful cases, the location of a gene could be specified within several million bases of DNA.

In the late 1970s and early 1980s, scientists took the first steps toward maps of human chromosomes based on direct biochemical analysis of

Figure 1-1 .—Comparative Scale of Mapping



The number of base pairs of DNA in human cells is roughly comparable to the number of people on Earth. The scale of genetic mapping efforts can be compared to population maps, with chromosomes (50 to 250 million base pairs) analogous to nations, and genes (thousands to millions of base pairs) to towns.

DNA. DNA fragments of unknown function but known location were used to study inheritance of traits far more precisely than before. Calculations suggested that DNA markers, which signify the presence or absence of particular stretches of DNA, could be identified for regions of all the human chromosomes.

Markers can be used to trace which pieces of DNA, and therefore which parts of chromosomes, are inherited from which parent. When a genetic trait is caused by a single gene and that gene is close to a marker, the marker can be used to ascertain roughly where the gene is located because the two are inherited together.

The U.S. Government and research agencies abroad fund most research that uses DNA markers to study diseases and physiological functions and most university groups searching for new markers in chromosomal regions of particular interest. In the United States, the National Institutes of Health (NIH) are the largest funding sources for biomedical research on genetics.

Construction of maps of DNA markers was undertaken in the early 1980s, The two largest collections of markers were developed by the Howard Hughes Medical Institute (HHMI), a private philanthropy, and Collaborative Research, Inc., a private corporation. Dozens of university researchers and other private firms also contributed to this kind of genetic map.

In 1985, DOE began planning the Human Genome Initiative to develop research tools for molecular genetics. Events leading up to the initiative included a workshop convened by the University of California at Santa Cruz and internal planning by DOE administrators. DOE considered the initiative an extension of its ongoing work in molecular biology-largely focused on detecting mutations and other biological effects of radiation and energy production-that would take advantage of research staff and instruments located at the national laboratories, which are funded by DOE, DOE held several public meetings to discuss the technical possibilities. The first of these was a workshop held in March 1986 in Santa Fe, New Mexico.

Discussion at that workshop of whether to establish a reference sequence for the entire human genome touched off a controversy that has persisted ever since. Arguments about the usefulness of extensive sequence information reached a high pitch at a conference at Cold Spring Harbor Laboratories in June 1986. Many scientists perceived a major sequencing effort as a threat to the conduct of basic research in molecular biology because of its projected cost and potential drain on research talent. Estimates of the cost of sequencing alone (without accounting for mapping or preparation of DNA to be sequenced) ran to billions of dollars. Calls for central management of such a prodigious undertaking further heightened tension because of the strong tradition of decentralized, small-group research in molecular biology. Debate over the appropriate strategy for deciding which regions to sequence first added to the din and spilled over into the scientific press. Major newspapers and magazines have covered the debate since, giving the Human Genome Initiative a high public profile,

The Cold Spring Harbor discussion was followed by a series of meetings held by HHMI, NIH, DOE, NRC, OTA, and others, Plans for special research initiatives by NIH, DOE, and HHMI have resulted from these and other discussions. A few private corporations have also been established (or are being established) to perform DNA sequencing and to develop research resources.

This report deals with various projects that have been proposed by Federal agencies to construct maps of human and other chromosomes, to improve relevant databases and repositories, and to improve research methods and instruments. **There is no single human genome project, but instead many projects.** For 1988, there are specific line items in appropriations for DOE and NIH, and the bulk of the discussion in this report refers to these new research programs. For purposes of this report, genome projects **refers to the research programs of NIH, DOE, and HHMI, as well as parallel programs in the private sector or other nations.**

THE FOCUS OF GENOME PROJECTS

Genome projects have several objectives:

- to establish, maintain, and enhance databases containing information about DNA sequences, location of DNA markers and genes, function of identified genes, and other related information;
- to create maps of human chromosomes consisting of DNA markers that would permit scientists to locate genes quickly;
- to create repositories of research materials, including ordered sets of DNA fragments that fully represent DNA in the human chromosomes;
- to develop new instruments for analyzing DNA;
- to develop new ways to analyze DNA, including biochemical and physical techniques and computational methods;
- to develop similar resources for other organisms that would facilitate biomedical research; and possibly
- to determine the DNA sequence of a large fraction of the human genome and that of other organisms.

Genome projects underway or planned by DOE, NIH, the National Science Foundation (NSF), HHMI, and other organizations are different but overlapping. They share two features: They would put new methods and instruments into the tool kit of molecular biology, and they would build a research infrastructure for geneticists (see table 1-1).

DOE's Human Genome Initiative began in late 1986 and consists of several projects. One is to create an ordered set of DNA segments from known chromosomal locations; this set, if widely available, could save the tedious steps involved in isolating DNA for study once a gene's approximate location is known. It should also reduce needless duplication of effort by different groups studying genes in the same chromosomal region. A second project is to develop new computational methods to enhance analysis of genetic map and DNA sequence data. Another project is to develop new techniques and instruments for detecting and analyzing DNA, including automation and robotics. For these projects, DOE expended $4.2 million in 1987 and plans $12 million for 1988. It also planned to support an additional $7 million in 1987

Table 1-1—Principal Organizations Involved in Genome Projects

Organization	Mission	Funding ($OOO,OOOs)[a]
National Institutes of Health (Department of Health and Human Services)	Biomedical research	Life sciences: 6,170 Related research: 313 Genome projects: 17.2 NLM biotechnology databases: 3.83
Department of Energy (Office of Health and Environmental Research, Office of Energy Research)	Biological effects of energy production and radiation; use of national laboratory resources	Life sciences: 230 Related research: 7 Genome projects: *12*
National Science Foundation (Directorate of Biological, Behavioral, and Social Sciences	Basic scientific research	Life sciences: 206 Related research: 32.7 Genome projects: 0.2
Howard Hughes Medical Institute	Biomedical research	Life sciences: 240 Genetics: 40 Genetic marker maps: 2 to 4 Databases: 2

[a]"Life sciences" figures are estimates for fiscal year 1987; these are total budgets for NIH and HHMI and estimates of relevant programs for NSF and DOE. Figures for "related research" include basic research projects that Involv mapping or sequencing, and research infrastructure such as databases and repositories. "Related research" figures are estimates for fiscal year 198?; "Genome Projects" figures are estimates for fiscal Year 1988, based on appropriations under the December 1987 continuing resolution.

SOURCES: NIH: Rachel Levinson, personal communications, October, November, December 1887, January 1888; DOE: David Smith, personal communications, June, October 1987, January 1988; NSF: David Kingsbury, personal communications, June and November 1987; HHMI: George Cahill, personal communication, January 1988.

for related research and infrastructure. DOE has requested $18.5 million for direct support of its Human Genome Initiative in fiscal year 1989.

NIH has supported special genome projects since 1987, with two objectives: to improve methods for analyzing the genome of human beings and other complex organisms and to enhance computational methods. NIH also supports most of the relevant databases and repositories. It spent an estimated $313 million on projects that involved mapping and sequencing in 1987, and several million more on infrastructure. NIH plans somewhat higher spending for related research in 1988 and will have two items in its budget—an additional $17.2 million for genome projects and $3.83 million for increased database support at the National Library of Medicine. The fiscal year 1989 budget request for the National Institute of General Medical Sciences of NIH includes $28 million for genome projects.

HHMI has two genome initiatives: one to support key databases containing information about the genetics of human and other organisms, and the other to support biomedical research on basic genetic mechanisms and genetic disease. HHMI's budget estimates from 1987 included $40 million for genetics (including $2 to $4 million for genetic mapping) and $2 million to support genome databases.

The NSF plans to increase the number of biology centers it supports, in order to develop new scientific instrumentation and encourage sharing of expensive equipment. These and other NSF programs are not genome projects per se, but they are likely to be integrated with programs of other agencies in some locations. Instrumentation developed through the biology centers will probably be directly relevant to genome projects. NSF budget estimates for 1987 were $206 million for life sciences, of which $32.7 million went to research related to genome projects and $200,000 went directly to genome projects.

Mechanisms for interagency coordination of genome projects have evolved over the past 2 years. Initially, there was informal communication among DOE, NIH, NSF, and HHMI. The Federal agencies then formed a working group under the Domestic Policy Council (DPC), a cabinet-level group in the White House. A committee to replace the DPC group is now being organized by the White House Office of Science and Technology Policy (OSTP), but its exact composition and function have not yet been determined.

International efforts are concentrated in developed nations with strong research traditions, Mapping genes, both human and nonhuman, has been an international effort since its inception. International agreements for databases (particularly those containing DNA sequence data) and collaborations on gene mapping (notably, the Center for the Study of Human Polymorphism in Paris) have been in operation for several years. No **foreign government has made a commitment yet to mapping and sequencing the human genome,** although several governments support related projects through their usual mechanisms of research funding. The United Kingdom has supported one of the pioneering efforts to map the genome of a nonhuman organism and additional work to develop new mapping and sequencing technologies. Italy has the most specific commitment to the human genome: It funded several pilot projects (up to $1 million per year for 2 years) to map and perhaps sequence at least one small human chromosome, with the intent of increasing that budget five- to ten-fold if the projects are promising. France, the Federal Republic of Germany, and other Western European nations have substantial commitments to genetics research and are also discussing international cooperation. Canada's medical research planning board is considering special efforts for genome projects. The European Molecular Biology Laboratory and European Molecular Biology Organization have expressed interest in an international collaboration to map and sequence the genomes of human and nonhuman organisms.

Eastern European and Asian nations have expressed interest in using the resulting data, but they have relatively limited programs for genetics research. Australia is one possible exception; it has consistently increased its share of publications related to genetics over the last decade, and it would logically be included in any international planning. Japan is another exception. Its Science and Technology Agency has expended $3.8 mil-

lion to support automation of DNA sequencing technologies, the Ministry of Education supports a grants program in genetics, and the Ministry of International Trade and Industry has devoted several million dollars to study the feasibility of an expanded international effort called the Human Frontiers Science Program, which could inelude genome research projects.

MISPLACED CONTROVERSY ABOUT "THE HUMAN GENOME PROJECT"

Over the past several years, the debate about genome projects has been vigorous–sometimes acrimonious. Many articles have appeared in the scientific press and the general press about "the genome controversy." The most conspicuous disagreements, however, have concentrated on issues that are not central to the conduct of genome projects, Disputes among the executive agencies have been played up, belying the generally close cooperation among DOE, NIH, HHMI, private firms, and other groups in conducting their respective projects. International cooperation among gene mappers and database managers has been successful but has attracted little attention. Private corporations are already involved in many of the projects that are furthest along. One firm has developed an extensive map of human genetic markers, and others have developed instrumentation useful in research relevant to the mapping and sequencing of DNA. These companies have offered few complaints about barriers to technology transfer. **Dissent has focused on the importance of and strategy for sequencing DNA of the entire genome, yet no agency has made a commitment to massive sequencing.** The current commitment is to develop technologies that would make it faster and less costly and to improve databases to collect and disseminate the resulting information. DOE has expressed interest in a concerted sequencing program, but only when technological development reduces its cost to tens of millions of dollars, in several years at the earliest.

Some of the debate can be attributed to the title that has often been applied to genome projects —the Human Genome Project. The term is a useful way to link research initiatives and to distinguish them from ongoing programs for budget planning. It highlights the *ultimate* objective– understanding human biology by developing a new set of research resources—and captures political support and broad public interest. It has had the effect, however, of generating rancorous debate which has inhibited the development of consensus on how to improve the research infrastructure. The importance of maps, databases, and repositories has been obscured by the controversy over massive DNA sequencing.

The title has had several other untoward effects. The Human Genome Project centers attention exclusively on human genetics, but **understanding human genes will necessarily involve the study of other organisms.** Many of the resources—particularly maps of human chromosomes—will be focused on human beings; but to interpret human genetic information, similar resources must be developed for other organisms. New instruments and methods will be applicable to all DNA.

The Human Genome Project invites confusion by implying that the human genome will be understood when the project is over. The immediate goal of genome projects is not complete understanding, but creating tools to bring about such understanding in the 21st century. Understanding encompasses all biomedical research; it does not distinguish genome projects from others. The most ambitious possible goal of genome projects would be to complete the most detailed map: a reference sequence of the entire human genome. Even if this were agreed to and developed, it would not yield immediate understanding of how that DNA sequence is translated to make a human being. It would not explain how nerve cells become connected in the immensely complex anatomy of the brain. It would not even provide complete answers to how individuals differ or how they have evolved. Sequence data, like other genetic information, is meaningful only when compared among individuals and correlated with biological function.

There is no single, monolithic Human Genome Project. In fact, there are several distinct components at various stages of development. Some instruments and many databases already exist; some genetic maps are more than half complete; repositories for DNA used in research have only been organized in the past few years; and other projects are planned but not yet begun. Whether there will ever be large and expensive research facilities for component specific genome projects is an open question that can be answered only as the technologies evolve.

The Human Genome Project conjures up images of largescale projects such as the Man= hattan Project to build the first atomic bomb, the Apollo Project for a manned Moon landing, the space station, or the superconducting supercollider. Genome projects do not belong in this category. Component genome projects will not require budgets as large as such megaprojects, nor are the technical ends as focused. Genome projects must be distinguished from the sequencing of the entire human genome, which is but a component still in the planning phase, There will be no single event such as the Moon landing or the space shuttle launching, nor is there likely to be construction of a new multi-billion-dollar facility such as the superconducting supercollider. Genome projects do not now require such facilities. Some projects may require facilities to perform services for mapping or sequencing in the future, yet such facilities would not be larger than the molecular biology centers already established at a few major research universities, Mapping or sequencing facilities would differ only by being devoted to production work rather than pure science. The results of genome projects are not contingent on completion of large capital-intensive dedicated units, and the data and instruments will be integrated into biology and medicine as the projects progress. Genome projects are, in this respect, analogous to navigational charts or road maps, which are useful even as they are being updated. Some persons believe a shortage of trained scientific and technical personnel in the United States could prove troublesome for molecular biology, but the genome projects proposed thus far are not so large in scale, even in comparison to other areas of biology, as to cause shortages in other areas. Genome projects are relatively modest compared to other large science projects now under consideration by the Federal Government,

THE CORE ISSUE: RESOURCE ALLOCATION FOR RESEARCH INFRASTRUCTURE

Most issues that need to be addressed regarding genome projects are variations on the problem of the commons: how to create and maintain resources of use to all. It can be difficult to develop goods useful to all if each individual has no direct incentive to pay for them and only a few are adversely affected.

The core issue concerning genome projects is resource allocation. What priority should be given to funding databases, materials repositories, genetic map projects, and development of new technologies? Should genome projects have precedence over other projects important to biological and biomedical research? These projects will benefit the entire biomedical research community, and ultimately the Nation and the world, but their funding must be drawn from the same agencies that support basic research. Funding for genome projects will thus be taken from agencies that support research on neuroscience, cancer, immunology, and many other promising and rapidly moving fields.

The flow of information from molecular biology is overwhelming the resources devoted to handling it. Federal agencies, HHMI, and other interested groups are acting to manage the deluge. **Research dedicated to improving databases, maps, repositories, and research methods premises to increase efficiency overall by de ing once systematically what would otherwise be duplicated by many groups using more primitive technologies.** Whether massive, concerted DNA sequencing is similarly efficient can only be demonstrated by trying it on a smaller scale.

ORGANIZATION OF THIS REPORT

The following sections describe options for congressional action, Subsequent chapters address the issues raised here in greater detail. Chapter 2 provides technical background and explains how genome projects might be conducted. Chapter 3 reviews how results might be used in biology and medicine. Chapter 4 outlines some long-term social and ethical issues surrounding human genome projects. Chapter 5 surveys agencies and organizations in the United States actively supporting human genome projects. Chapter 6 discusses how genome projects might be organized among these agencies and organizations. Chapter 7 briefly surveys activities in foreign countries, and chapter 8 presents issues involved in technology transfer, Appendixes contain background on material used to produce this report, databases, costs of projects, and mapping and sequencing publications.

THE ROLE OF CONGRESS

Genome projects have come to the attention of Congress for three reasons. First, they have become highly visible because of the extensive debate surrounding them. Second, they involve agencies in different executive departments; therefore, mechanisms for coordinating them are less clear than if they were all in a single department. Third, results of genome projects will lead to new scientific and medical instruments for analysis of DNA, development of new genetic tests for use in clinical diagnosis, and other products and services, Techniques developed to analyze DNA will expedite biological research and will provide data and technologies crucial to the development of many new products. In this sense, genome projects promise economic returns, although the form and magnitude of them are not predictable. Genome projects have thus been linked to international competitiveness in biotechnology and its economic implications for American commerce in coming decades.

Congress has three roles regarding genome projects:

1. annual *appropriations* to Federal agencies funding the projects;
2. *authorization* of actions by executive agencies to setup formal coordinating structures or of specific mandates of agencies; and
3. *oversight* of agencies' conduct of their projects.

OPTIONS FOR ACTION BY CONGRESS

Options for congressional action discussed here build on the discussions above and those in chaptees 4 through 6. Background material and details can be found in those chapters.

Appropriations to Federal Agencies

The pace of federally funded genome projects will be determined principally by the annual appropriations set by Congress and by the executive agencies' commitment to the projects, Although agencies retain some authority to "reprogram" funds for activities that fall within their mandates, large efforts cannot be sustained without specific appropriations. Appropriations will set an upper limit on the size and number of projects that are federally supported; commitmerit by executive agencies, and their grantees and contractors, will determine the speed and scope of projects within those limits.

The critical judgment in appropriations is the importance of the work to be supported relative to other research and activities supported by the Federal Government. The two national scientific groups that have written reports on genome projects, a DOE advisory subcommittee and an NRC committee, have both recommended substantial additional funding for genome projects, eventually equaling $200 million per year. OTA inde-

pendently projected costs of genome projects at a workshop and through subsequent interviews and letters. Appendix B summarizes cost estimates, including the history of those made by other groups, and reviews the process OTA used to make its estimates. The cost of funding all component projects was estimated as increasing from $47 million the first year to $228 million the fifth year. This would permit strengthening of databases and repositories, construction of several varieties of chromosomal maps, development of many new technologies, and initiation of pilot projects for DNA sequencing.

Access to Information and Materials

The information produced by genetics research has swamped existing management systems. Materials to facilitate molecular genetic research have also proliferated, straining the resources devoted to making them widely available. These management problems will intensify as new technologies further accelerate research. Several of the genome projects are intended to systematically archive information, collect and store research materials, and make information and materials widely available to the research community. **Improving data= base and repository services is imperative whether or not other genome projects pm ceed.** If genetic mapping and sequencing initiatives are pursued, then databases and repositories will be needed even more. Bills have been introduced to improve coordination of and access to molecular biology databases through the National Library of Medicine. Each major repository and database has its own advisory panel of outside scientists. NIH has appointed an internal committee to report on NIH-supported repositories. Two international meetings were held in 1987 to discuss management of databases that contain DNA sequence data. NIH and DOE cosponsored a meeting on databases and repositories in August 1987, and appropriations to DOE and NIH have been increased to support databases and repositories. Congress has the options of maintaining current funding levels or increasing funds for database and repository services through the current system of agency planning and congressional oversight. Seeking recommendations from an advisory committee on how to integrate the development of databases and repositories with genome projects is an additional option.

Organization of Genome Projects

Congress could pass legislation to organize human genome projects—in fact, bills on organization have dominated discussion in Congress. There are four principal choices: 1) to designate a single agency to coordinate the projects, 2) to establish an interagency task force, 3) to establish a national consortium, or 4) to rely on congressional oversight of interagency agreement and consultation.

Establishing an interagency task force through legislation or encouraging agencies to do so by oversight are the least problematic choices. Designating a lead agency would be politically troublesome and would risk interruption of ongoing research programs atone or more agencies. Devising a single national consortium to manage the many diverse genome projects is likely.to prove impractical. See chapter 6 for a more detailed discussion of these options.

Designate a Lead Agency

Congress could choose to designate a lead agency to coordinate and provide principal funding for genome projects. The chief advantage of a lead agency is accountability through clear authority. The purpose of focusing authority would be to reduce duplication of effort, to enhance coordination, and to give Congress a single agency on which to concentrate oversight. The chief disadvantage is that the difficult political process of selecting a lead agency would delay progress and diminish overall funding. If line item funding for genome projects at the nonlead agency–NIH or DOE—were eliminated, then agreement would have to be reached to add funds for the lead agency. This is a difficult process because it involves a completely different set of congressional committees and subcommittees for each agency. The choice of a lead agency would likely precipitate a protracted battle among agencies and congressional committees, which could only serve to delay projects. Furthermore, activities of NIH, DOE, NSF, HHMI, and other organizations are complementary rather than competitive and duplicative. Appointing a Iead agency could complicate

planning for the other agencies. As an alternative, each agency could take the lead in projects best suited to its mandate and expertise. This would result in a task force or consultative arrangement, discussed below, rather than a single lead agency. Designating a lead agency would attempt to centralize authority, but it is not clear that this would improve efficiency, communication, or coordination.

Designation of a lead agency for genome projects could, paradoxically, diminish rather than enhance accountability to Congress. This follows from the organizational structure of congressional committees. Genome projects supported by NIH, DOE, and NSF are authorized by several committees and subcommittees in both the House of Representatives and the Senate. Currently, each committee or subcommittee has independent authority to oversee programs in agencies under its jurisdiction, and interest in human genome projects has been high. Designating a lead agency would limit most oversight responsibility to a single committee. Further, a lead agency could not fully centralize authority, because HHMI is a nongovernment organization. Picking a lead agency would be politically difficult and is unlikely to occur unless there is strong evidence of the advantages of centralized authority for Federal efforts. The evidence to date is quite to the contrary: Agencies are communicating, sharing personnel, using compatible peer review procedures, and jointly funding projects in overlapping areas.

Designating a lead agency might eliminate pluralism in Federal funding of genome projects. An investigator wishing to pursue a genome project can now apply to NIH and DOE, or NIH and NSF for funding (depending on the nature of the project), If there were a single lead agency controlling genome projects, the choices would be limited, diminishing the pluralistic funding that has been a mainstay of American biology. If the lead agency had only an administrative role and did not provide the greatest amount of funds, then there would be little point in calling it a lead agency.

Congress sets independent budgets for NSF, NIH, and DOE through different subcommittees in the House and Senate appropriations committees. With several subcommittees involved, projects have alternative sources of support in Congress. Designating a lead agency would reduce this flexibility. The danger of pluralism is that different agencies will duplicate each other's work, will fail to cooperate, will fail to identify gaps in research, or will receive uncoordinated or inappropriate appropriations due to the absence of a clear authority structure. To date, such funding disarray has failed to materialize. There are checks and balances in the congressional budget process, through the Office of Management and Budget (OMB), and through the interagency consultation group in OSTP.

Arguments for a centralized and highly organized effort would be stronger if genome projects addressed a national health emergency, such as AIDS or polio, or if they were aimed at a single technical or scientific objective. But genome projects are many and diverse. Focused responsibility may nonetheless become necessary for some of them. Mapping, for example, might be more efficiently done at production centers as methods mature, and DNA sequencing might require dedicated facilities if the technology demands high capital investment or central management, If dedicated service centers are established, administration by a single agency or formal interagency agreement would be necessary to ensure standardization and efficiency. Such services would only be components of overall ge - nome projects, however; integration of the various projects would still be needed.

If genome projects were neglected or inconspicuous elements in agencies' programs, then the advantages of central oversight through a single agency would carry more weight. This has not been the case. Genome projects have been given high priority–first by DOE and more recently by NIH—and there has been extensive media attention to agencies' management of them. There is thus little danger in the foreseeable future that genome projects will receive insufficient attention or that mismanagement will escape congressional scrutiny.

The agency most affected by genome projects will be the NIH. If Congress finds that the advantages of a lead agency outweigh the disadvantages,

then NIH is the natural choice for lead agency. This is because biomedical research is NIH's central mandate, whereas NSF's and DOE's research programs include physical as well as life sciences. NIH funds over 10 times more genetics research than any other government or nongovernment organization, and researchers funded by NIH are the most numerous of the intended beneficiaries of genome projects. Researchers supported by DOE, NSF, and other organizations have important contributions to make, however, and some projects fall outside the mainstream of research supported by NIH. Genome projects that involve expertise in physical science, engineering, and other fields outside biomedical research would benefit from participation in or leadership by NSF or DOE. DOE in particular has vigorously promoted a Federal program to develop new technologies and to create sets of ordered DNA fragments. Some DOE-supported projects are logical extensions of work at the national laboratories, and DOE is the natural agency to conduct these. If NIH were designated the lead agency, it would be important to recognize and plan for the ongoing efforts of DOE.

Establish an Interagency Task Force

The chief advantage of an interagency task force is that it builds on existing research programs and planning efforts in different agencies and does not require a single lead agency. A task force could monitor all genome projects, government and nongovernment, obtain scientific advice, foster communication, and make recommendations to Congress and the appropriate agencies. Discussion at an OTA workshop in August 1987 stressed that agencies should have outside scientific advice and that advice given to one agency should take into account activities supported by other agencies. No advisory body exists to carry out this task. The chief disadvantage of a task force is that no one agency is accountable for the conduct of genome projects.

Creating a task force entails decisions about who should be represented, how appointments are to be made, and where the task force would be located administratively. Legislation could specify that it represent government, academic, industrial, and other relevant expertise and could stipulate the terms of membership and the appointment process. The task force could be made part of a government agency (making it in effect the lead agency), administratively autonomous, or attached to an existing quasi-governmental institution such as the National Academy of Sciences. Several bills to establish such coordination and advisory groups have been introduced in the looth Congress and are likely to be acted upon in 1988.

Create a National Consortium

A consortium would involve one or more agencies in concert with private partners to support genome projects. The chief advantages of a consortium are administrative flexibility, possible funding by private firms to reduce government funding, and direct involvement of industrial partners—which would presumably hasten technology transfer. Some potential disadvantages are unclear lines of authority (caused by competing needs of government and private partners) and statements by the private sector that genome projects should be funded exclusively by the Federal Government (e.g., a poll taken by the Industrial Biotechnology Association). Accountability would be complicated in two respects. First, there are many genome projects, and it is difficult to imagine a single consortium that could oversee them all. Second, the possible commingling of government and nongovernment funds could prove troublesome. Consortia might nonetheless be formed for specific tasks. Some genome projects in technology development will undoubtedly be of great interest to industry and might attract private funding. Such projects (e.g., developing automated DNA mapping instruments or DNA detection methods) are likely to be highly focused, however, and organized at the local rather than the national level. Accountability would not be as diffuse for local consortia focused on specific technical objectives as for a single national consortium with multiple objectives and dozens of projects to manage.

The Technology Transfer Act of 1986 (Public Law 99-502) grants government agencies authority to form consortia with private corporations and provides guidelines for doing so. President Reagan's Executive Order 12591 (April 1987) further extends this authority and encourages fed-

erally owned laboratories to form consortia. Agencies thus have the requisite authority already. If Congress finds terms of the 1986 bill inappropriate in some details—for example, regarding patent policies or royalty arrangements—then the statute could be amended or special measures relating to genome projects could be added as amendments to other bills.

One bill introduced early in the 100th Congress would have established a national consortium specifically to manage genome projects, but the bill has since been replaced by one that establishes a new advisory body (covered above as a task force). A national consortium is not the only, and perhaps not the most effective, way to obtain industrial input for genome projects and to facilitate technology transfer. Alternatives are to encourage agencies to participate in the formation of local consortia; to facilitate exchange of industrial and academic expertise through training exchange programs, symposiums, and other mechanisms; and to include industrial representation on any national advisory groups.

Rely on Congressional Oversight

If Congress takes no explicit action, several outcomes are possible. Federal agencies could continue planning processes similar to those followed in 1986 and 1987, consisting of informal communication and coordination through an interagency group with members from NIH, DOE, NSF, OSTP, OMB, and other agencies. To date, NIH, DOE, and NSF have sought outside advice from various standing advisory committees, a practice that has resulted in conflicting recommendations. This problem could be remedied without legislation: The agencies could establish a single interagency advisory committee of outside experts appointed by the agencies or by a third party, such as the National Academy of Sciences or a private philanthropy. The advisory committee could report to the agencies directly.

A Committee on Life Sciences is forming in OSTP. The interagency nature and conspicuousness of genome projects make them a natural topic for this committee. OSTP is considering the creation of a special subcommittee on genome projects. Whether OSTP'S efforts meet the objectives desired by Congress will depend on effective coordination and an appropriate balance among government, university, industrial, philanthropic, legal, bioethical, and other representatives on the subcommittee. If OSTP'S subcommittee is composed exclusively of government representatives, then its primary function will be interagency communication. The main stumbling block to interagency planning to date has been conflicting advice from outside advisory bodies, not lack of interagency communication. Pluralism in funding is usually a virtue, but making conflicting recommendations to different agencies is not. Any national coordinating group should take a global view of activities in all agencies and harmonize the advice given them.

The chief advantage of relying solely on congressional oversight is that it requires no new legislation. One disadvantage is that interagency agreement on appointments and operating budgets for a coordinating body might prove difficult without a congressional mandate and might not initially include an appropriate range of nongovernment experts. Another potential disadvantage is that initiatives undertaken by an administration in the absence of legislation could crumble under the weight of later interagency disagreements or neglect by a subsequent administration. Flexibility is beneficial if projects are short-lived, but genome projects are not. Long-term stability is essential to the efficient conduct of genome projects because they will require sustained support over many years. Oversight of agency action could nonetheless be all that is required. Deficiencies of a task force set up by agencies could later be modified indirectly through congressional oversight or threat of legislation.

Technology Transfer

Congress appropriates funds to support scientific research for several reasons, the principal one for biomedical research being to improve health. Increasingly, however, biomedical research is being regarded as a national investment, and policies to facilitate economically fruitful applications of new knowledge are receiving attention in Congress. The process of exploiting new knowl -

edge for practical purposes is called technolo~ *transfer.* Some persons favor increased funding for genome projects because they believe the projects will lead to marketable products (instruments, research materials) or will accelerate research in areas that will later yield marketable products. Technology transfer can be improved through patent policies, exchange of industrial and academic personnel, symposiums for industrial and academic scientists, formation of consortia to develop specific technologies or services, and engaging industry in planning genome projects. Programs for exchanging personnel and sponsoring symposiums will fall to agencies through normal policy paths and can be monitored by Congress. Consortium formation and industry representation on planning bodies have been discussed above. The remaining policy area is patent and copyright law.

Patent policies of Federal agencies have changed dramatically during the past decade. The Patent and Trademark Amendments of 1980 (Public Law 96-517), as amended in 1984 (Public Law 98-620), were devised to facilitate commercialization of federally sponsored research. President Reagan issued directives to Federal agencies *in* February 1983 and April 1987 to this same end. And Congress passed the Technology Transfer Act of 1986 (Public Law 99-502), which contains patent licensing and joint venture provisions with authority to form consortia with private interests. These patent policies, following outlines of policies pioneered by NIH and NSF in the late 1970s, encourage institutions receiving Federal grants or contracts to patent products and processes resulting from federally funded work. A 1987 General Accounting Office report judged that the policies have increased patenting of research results.

Aside from a possible change regarding DOE policies (see ch. 8), genome projects raise no new questions of patent or copyright law. Genome projects would be subject to the same statutes and executive orders as other scientific efforts. There is a clear role for congressional oversight, however, in ensuring that data are shared promptly and fully.

In mid-1987, proposals to form private corporations to map and sequence the human genome stirred a controversy. Scientists expressed concern that scientific exchange would be impeded by such efforts and that information would be sequestered through copyrights and patents. If private corporations do form to develop map and sequence data and research materials, they will operate at private expense. If they are successful, scientists will have new information, services, and materials available for a price. If they fail, scientists should be no worse off, unless the government fails to support work it would otherwise have funded. To date, government agencies have not dropped plans for genome projects because of corporate efforts.

Corporate efforts need not entail restricted access to information, Corporations can provide services not appropriately performed by laboratories conducting basic scientific research (e.g., mapping, sequencing, or database management). Universities and large corporations can manage research facilities, such as the national laboratories, under contract. Corporations could also participate in consortia focused on specific technical objectives. Private firms could be given grants to develop new methods under the Small Business Innovation Research program; they would retain title to inventions, but they would have the same obligation to share data and materials as universities or other grantees. The essential point is not whether a grantee or a contractor is a university or corporation, but whether the research results will be widely shared.

It is essential to ensure timely exchange of data and materials from federally sponsored projects. Maps, databases, and repositories will be useful only if they are accurate and complete; they will be complete only if all participants make prompt contributions. Inmost cases, patent requirements should not substantially delay disclosure of data. Many data will not be relevant to a patentable invention. When research results do include a patentable invention, advance planning for filing pat - ent applications should minimize delays. The main option for Congress in this area is to oversee the conduct of genome projects. Changes in agency policies for data exchange could be made if problems emerge.

Congress could also direct agencies to make it easier for persons receiving Federal grants or contracts to understand patent policies in the United States and abroad. At present, many of the published guidelines and regulations for NIH, DOE, and NSF are out of date. Investigators contemplating genome projects will probably contact more than one Federal agency for research support; it would be helpful to have a document summarizing the practices of the different agencies. Such a document could also explain the benefits of filing patents early and outline procedures for patenting abroad.

Questions for Congressional oversight

Congressional oversight will most often involve an informal exchange among congressional staff, executive agency personnel, and other interested parties, Oversight can be a potent incentive for cooperation among agencies and for good conduct of executive actions. Congress may wish to hold hearings from time to time to address such questions as: Are genome projects being efficient} administered? Are agencies duplicating efforts on genome projects? Are agencies communicating effectively? Are agencies ensuring that access to shared data is relatively easy and fair? Are databases receiving the information they need to be most useful (e.g., map and sequence data)? Are commercial opportunities being exploited? Are shared research resources being neglected? Are issues of special interest to Congress, such as social and ethical implications of genome projects, being adequately addressed? Do genome projects supported by Federal agencies reflect national needs and social priorities? Are foreign go/ern - ments funding a proportionate share of genetics research and the research infrastructure? Are foreign governments sharing data and materials to the same degree as LT.S. agencies?