28. J. D. Watson and E. Jordan, *Genomics* **5**, 654 (1989).
29. D. J. Kevles, *In the Name of Eugenics* (Univ. of California Press, Berkeley and Los Angeles, 1985); R. Proctor, *Racial Hygiene* (Harvard Univ. Press, Cambridge, MA, 1988).
30. Committee for Responsible Genetics, *Position Paper on Human Genome Initiative* (Committee for Responsible Genetics, Boston, MA, 1990).
31. B. Muller-Hill, *Murderous Science* (Oxford Univ. Press, New York, 1988).
32. J. Wyngaarden, personal communication.
33. J. Palca, *Nature* **337**, 200 (1989).
34. ———, *Science* **245**, 131 (1989).
35. C. McGourty, *Nature* **340**, 496 (1989).
36. L. Roberts, *Science* **241**, 1596 (1988).
37. ———, *ibid.* **247**, 281 (1990).
38. M. Olson, L. Hood, C. Cantor, D. Botstein, *ibid.* **245**, 1434 (1989).
39. STS is an acronym for sequence-tagged site. An STS is basically a short DNA sequence that has been shown to be unique within a genome.
40. Y. Kohara, K. Akiyama, K. Isono, *Cell* **50**, 495 (1987).
41. M. V. Olson *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **83**, 7826 (1986).
42. A. Coulson, J. Sulston, S. Brenner, J. Karn, *ibid.*, p. 7821.
43. A. Coulson, R. Waterston, J. Kiff, J. Sulston, Y. Kohara, *Nature* **335**, 184 (1988).
44. P. H. St George-Hyslop *et al.*, *Science* **235**, 885 (1987).
45. V. A. McKusick, *Genomics* **5**, 385 (1989).

Orchestrating the Human Genome Project

CHARLES R. CANTOR

The Human Genome Project is under way. The Department of Energy and the National Institutes of Health are cooperating effectively to develop organizational structures and scientific priorities that should keep the project on schedule and within its budget.

THE HUMAN GENOME PROJECT IS BIOLOGY'S FIRST LARGE science project with a definite end point. Although it is small compared to most other Big Science efforts, many biologists are still somewhat fearful of the impact this project will have on biology research traditions and funding priorities. Here I outline how the project has evolved from its earliest conceptions to the present, rather different structure. My intention is to convince the reader that a productive, sensible, scheme is in hand to manage this effort and to achieve the goals of the project within a reasonable budget and time period. The short-term cost to traditional biology should be small, but the long-term benefits should be almost unmeasurable.

The Human Genome Project appears to have had several independent origins. One started in a meeting in Alta, Utah, in 1984, when a number of scientists began thinking about the prospect of sequencing all the DNA in the human genome (*1*). The meeting was not called for this purpose. Under the auspices of the U.S. Department of Energy (DOE), Ray White and Mortimer Mendelsohn had convened a small group of experts, mostly molecular biologists, to try to solve a problem. The DOE has a congressional mandate to monitor inherited damage caused by low-level exposure to radiation and other environmental hazards. Existing methods simply were not capable of detecting mutation rates in exposed human populations. Tools were needed that could detect a single altered nucleic acid base in, say, 10^8. However, doing that would be almost as much work as sequencing the human genome.

Other significant origins of the Human Genome Project include a meeting organized by Robert Sinsheimer (*2*) at Santa Cruz in 1985 and an article by Renato Dulbecco (*3*) in 1986. All these roots seem to have coalesced for the first time at a meeting in Cold Spring Harbor in 1986 when the current model of the project as a multicenter, multinational cooperative effort reached full bloom.

More than 5 years after the first conceptualizations, we remain a long way from sequencing any complex genome, and even complete bacterial sequences have still been elusive. However, many initial skeptics have become convinced that mapping and ultimately sequencing the human genome and other complex genomes is a practical and worthwhile task. Our perspective of how to organize it has changed considerably, partly in response to concerns about the costs involved, concerns from the biological community, and changes in technology and strategies.

In the years immediately after the Alta meeting, a major stumbling block was finding people who would want to do such a seemingly boring and tedious task as sequencing the genome. Indeed, Sydney Brenner has jokingly suggested establishing a penal colony where sentences consisting of large-scale sequencing projects would be carried out (*4*). A popular model was a large center, highly integrated and organized along industrial lines. Walter Gilbert made a strong case that there was no reason for delay as the technology was in hand to do the project at a cost that would be dwarfed by the ultimate benefit (*5*). However, a majority of the early enthusiasts for the project felt that initial, major investments in improvements in technology would soon result in much more efficient gene mapping and sequencing methods. This would greatly increase the power of individual investigators and obviate the need for a massive central structure. This model, with evolving technologies playing a major role, fit in much better with the spirit of contemporary biological research, and it ultimately became the accepted framework. It carries the explicit assumption that the cost of DNA sequencing must be reduced by at least an order of magnitude before the major sequence production aspects of the final project can commence.

Several research developments helped stimulate broader interest in generating complete human genomic maps on a reasonably short time scale. The completion of physical maps of *Escherichia coli* showed the feasibility of such projects (*6*), and the immediate usefulness of these maps in a variety of biological experiments ranging from finding genes to characterizing DNA rearrangements made the project seem less onerous. Excitement was generated when several important human disease genes were located by a combination of genetic mapping and molecular biological analysis. Howev-

The author is the director of the Human Genome Center, Lawrence Berkeley Laboratory, Berkeley, CA 94720.

er, the cost of finding genes individually was quite inhibiting, and the efficiency of more global approaches was very attractive. DNA sequence data have continued to accumulate and, as sequences of different genes have been compared, several striking examples have been found whereby proteins with no prior hint of common function showed extensive sequence homology. For example, the *sis* oncogene appeared to be closely related to platelet-derived growth factor (7), whereas the β-adrenergic receptor and rhodopsin displayed evidence of a common heritage (although they have evolved into processors of very different types of signals) (*8*). These unexpected matches not only provided major biological insights, but also suggested that a complete human genome sequence would not just be a meaningless exercise in data collection. A significant portion of the sequence would be immediately interpretable by association with genes of known functions.

The DOE was primed to become interested in the genome project because of two ongoing activities at its national laboratories. The National Gene Library project, established in 1985, took advantage of the highly refined flow-sorting capabilities of both Los Alamos (LANL) and Lawrence Livermore (LLNL) national laboratories. In this project individual human chromosomes were purified in sufficient quantities to make single-chromosome libraries. The earliest libraries had some deficiencies, but as techniques improved, the resulting sets of clones became an extremely valuable resource for investigators worldwide who were interested in particular regions of the human genome. The other DOE activity was the establishment in 1983 of the major U.S. DNA sequence database, Genbank, at LANL. It became very clear that organized databases and advanced methods of data management and analysis would be critically important as our knowledge of DNA maps and sequence grew.

DOE's interest in the genome initiative received a major boost when Charles DeLisi became the director of the Office of Health and Environmental Research. It soon became clear that DOE wanted this project. It was prepared to commit some resources to it immediately, and it would strive to secure much larger resources in the future. The national laboratories were definitely interested and had great expertise in instrumentation and computer science that were needed in the project. A series of DOE-sponsored meetings on the genome generated increasing interest and publicity, and soon afterwards, committees were set up by both the National Research Council and the Office of Technology Assessment to study the feasibility and desirability of mapping and sequencing the entire human genome. The model of the Human Genome Project that emerged from both of these studies was quite different from the original concept of crash programs of data collection in a few large production centers. A heavy initial investment would be made in improving technology. Much of the genome would ultimately be mapped, and virtually all of it sequenced, by methods that did not yet exist. There was no need to carry this work out at a single location, but the demands of new instrumentation and informatics needs suggested that much of it should be focused at a set of research centers created for this task. However, the desirability of the project was clear, and it was deemed sufficiently important that support was recommended at a large enough scale to allow the project to be completed in a relatively short time, 10 to 20 years (*9*).

A major additional theme addressed by both committees was the role that should be played by DOE versus that of the National Institutes of Health (NIH). DOE appeared willing to do the entire project if NIH was not interested. This raised a number of legitimate concerns in the biology community. Clearly NIH ought to be involved in the genome project. Its grantees would be the major users of the data to be generated. Most of the research in human genetics, and the DNA methodology that led up to the genome project, had been supported by NIH funds through ordinary research grants. However, at several levels, the initial reaction of NIH to the Human Genome Initiative was less than enthusiastic. On the scale of the typical NIH grants, the costs of genome mapping-sequencing prospects seemed large indeed. Furthermore, NIH review committees were accustomed to research projects based largely on the testing of a particular hypothesis. They had shown themselves to be reluctant to fund the large data-gathering exercises that would be typical of genome projects, and they were usually not eager to support the development of instruments or data-handling capabilities. For all these reasons it took NIH quite a while to become fully committed, and then only after it was decided to set up specific new structures at NIH to administer the project.

Congress is currently funding genome initiatives at both NIH and DOE. The two agencies have different skills and outlooks on the project that complement each other nicely. NIH must respond to the needs of the broad community of biologists. Thus it has chosen to expend considerable funds on refinements in human gene mapping because of its focus on human diseases, and on studies of the genomes of selected model organisms. The genomes of nonhuman organisms will be mapped and sequenced as a testing ground for determining gene function. The most likely way of testing the function of new human genes as they are found will be to find and manipulate equivalent genes in model systems. Almost all of the initial NIH support has been dispersed in traditional research grants, evaluated and managed in the manner standard at NIH. As the genome project matures and enters more production-oriented phases, it seems certain that NIH will focus an increasingly larger share of its support in a set of newly created, specialized centers.

DOE needs the results of the genome project, both the DNA sequence and the new methods used to obtain it efficiently, to carry out its mandate to monitor the inherited effects of low-level exposure to radiation and other environmental hazards. DOE has focused its initial attention solely on the human physical map and on the development of new methods of mapping, sequencing, and managing the data generated by the project. DOE is comfortable in managing large, long-term, applied science projects, and coordination of DOE efforts should be relatively easy since most of its current support is centered at just three national laboratories. By building on existing expertise in the national laboratories, DOE is in a good position to implement new programs quickly in such areas as informatics, robotics, and instrument development. The national laboratories, with long-term or permanent scientific staff and the infrastructure needed to manage targeted research, seem particularly appropriate places to carry out such tasks as large-scale mapping and sequencing and to handle the inevitable demands for samples, data, and follow-up studies.

By working together on the genome project, NIH and DOE have much to gain from each other. The current split of cooperation between the two separate programs is excellent. Representatives of NIH and DOE programs, supplemented by other individuals, have written a joint plan for the first 5 years of the project (*10*). In addition to coordinating various individually funded efforts, NIH and DOE have merged their activities in a number of key areas where separate parallel arrangements would be inefficient or distracting. These include a joint task force to oversee the challenging informatics needs of human genome analysis and a joint committee to deal with the significant ethical problems that will accompany the success of the genome project. Among the key informatics issues are (i) the need to access numerous databases in a way that does not preclude a user's particular software and hardware configuration, (ii) the desire to have real-time, remote entry of data, and (iii) the ability to correct historical records and thus eliminate the most serious potential causes of confusion in such multiuser situations. Among the likely key ethical issues are the need to ensure privacy of the

information gained about individual genomes in the context of demands by employers and insurers, and the potential discomfort to individuals who have knowledge of their genetic disease predispositions in situations where therapeutic measures are unknown. The issues are not simple ones, but pooling all of our efforts is a step on the right path to dealing with them.

At present, NIH-DOE coordination is occurring at three levels. Administrators from the two agencies are meeting regularly, dealing with funding, organizational, and prioritization issues. A joint subcommittee made up of members of the two genome program's advisory committees has been established to implement the Memorandum of Understanding signed by both agencies in October 1988 (*11*). It has been meeting quarterly to provide overall scientific coordination. Finally, individual scientists supported by one agency or the other are, as usual, unconstrained by their source of support, and numerous collaborative efforts are a hallmark of current genome research efforts.

How will all this be integrated with activities going on elsewhere in the world? The answer depends in large part on how quickly other national or multinational groups develop. Obviously, all humanity is represented in the human genome and in this project. However, the need to obtain the human gene map and sequence data is too urgent to wait for complex organizational or other political issues to be resolved. Traditional human genetic research has successfully revealed the identity of a number of important human disease genes. However, the cost of finding individual disease genes, one at a time, is often staggering. The ultimate benefits of finding even one major disease gene that might not have been observed by methods less systematic than the genome project could well recoup the entire cost of that project. Thus, while welcoming international collaboration and cooperation at all levels, it seems prudent for U.S. efforts to press ahead as the mechanisms of international cooperation become better elaborated.

In the long run it appears to make most sense to organize the bulk of the Human Genome Project around activities on individual chromosomes. Although it is unlikely that these can be effectively parceled out as assignments, it is very likely that individual interest groups will evolve into effective working partnerships for each genome region. This is balkanization in a positive sense because each group will share an organized set of materials, discover a unique set of DNA genes and probes, and share in the satisfaction when its part of the genome project is completed. There is room in the context of the Human Genome Project for both large and small scientific efforts. Much of the detailed mapping and sequencing is best done at fairly large centers, but more focused smaller efforts can easily be integrated into these programs. Technology development should be carried out in a variety of environments. At present DOE is devoting about two-thirds of its funding to large centers, while the remainder is spread among many locations through ordinary peer-reviewed research grants.

It is interesting and instructive to compare the Human Genome Project in its current stage and form with other large science and engineering efforts such as putting a man on the moon or various astronomy or high-energy physics projects. Like these other projects, the Human Genome Initiative has defined achievable goals. Major improvements in technology have been required in all of these projects, and much of the first half of the Human Genome Initiative will be devoted to attaining the desired efficiency. The technological innovations will include new mapping, sequencing, and informatics methods that will have an impact on biology far beyond the Human Genome Project itself, just as other large science projects have generated new techniques. Probably the largest impact will appear in an increasing use of computers by biologists. From origins such as the Genbank data repository we are likely to witness a great expansion in biological databases and networking.

Just as the construction of a telescope or an accelerator provides a tool for astronomers or experimental physicists, the human genome map and sequence will be tools that are usable by a broad class of biologists. However, there are some interesting differences between the impact of the genome project and other typical large science projects on the ordinary researcher whose projects are not directly related to the genome project. First, the tools will be accessible to all biologists and will be usable in the style and context of ordinary small science research. Second, the tools are guaranteed to be useful to a large number of biological researchers. Each human gene found will provide the raw material for a scientific lifetime of biological research. For those not interested in human biology directly, the human gene sequences will still be an important resource to uncover comparable animal or plant genes.

What will happen after the completion of the genome project? Will we be left with a host of unneeded personnel and equipment? The answer is an emphatic no. As part of the project, individuals will be trained in interdisciplinary areas such as biological instrumentation and bioinformatics. Such individuals will be a valuable future resource, and it is likely the genome project will serve as a major stimulus to create a new academic discipline in applied biology, just as large physics projects have stimulated applied science areas such as accelerator physics. The Human Genome Project is sure to stimulate the organization of other large genome projects. Many of these will center on animals and plants of commercial importance. These projects will probably be done in the private sector, although the Department of Agriculture plans to stimulate some early efforts in universities and research institutes (*12*). Thus the potential impact of the Human Genome Project on the biotechnology industry will extend far beyond its medical applications.

The past few decades have been very exciting for biology. Now on the threshold of its first large organized project, it is certain that biologists in particular and humanity in general will obtain remarkable benefits over the next 15 years, with relatively insignificant risk. Advances in technology even in the first year of the project have been more rapid than anticipated in early first estimates for DNA sequencing and associated activities. This makes me quite optimistic that the genome project can be completed on schedule and within its budget. It is both a pleasure and privilege to be involved in this effort.

REFERENCES AND NOTES

1. R. M. Cook-Deegan, *Genomics* **5**, 661 (1989).
2. R. L. Sinsheimer, *ibid.*, p. 954.
3. R. Dulbecco, *Science* **231**, 1055 (1986).
4. S. Brenner, Human Genome I conference, sponsored by *Science* and the Human Genome Organization, San Diego, CA, 2 to 4 October 1989.
5. W. Gilbert, *Basic Life Sci.* **46**, 29 (1987).
6. C. L. Smith, J. G. Econome, A. Schutt, S. Klco, C. R. Cantor, *Science* **236**, 1448 (1987); Y. Kohara, K. Akiyama, K. Isono, *Cell* **50**, 495 (1987).
7. R. F. Doolittle *et al.*, *Science* **221**, 275 (1983).
8. R. A. F. Dixon *et al.*, *Nature* **321**, 75 (1986).
9. National Research Council, *Mapping and Sequencing the Human Genome* (National Academy Press, Washington, DC, 1988); Office of Technology Assessment, U.S. Congress, *Mapping Our Genes—The Genome Projects: How Big, How Fast* (OTA-BA-373, U.S. Government Printing Office, Washington, DC, 1988).
10. L. Roberts, *Science* **245**, 1438 (1989).
11. ———, *ibid.* **241**, 1596 (1988).
12. C. Strange, *BioScience* **39**, 760 (1989).

13. Supported by Outstanding Investigator Grant CA 39782 from the National Cancer Institute, administered by the Lawrence Berkeley Laboratory, and by the director, Office of Health and Environmental Research, Human Genome Program, of the DOE under contract DE-AC03-76F00098. I thank S. J. Spengler for assistance and advice during preparation of this manuscript.