selves. Fermilab's fixed-target complex consists of three main areas: meson, neutrino, and proton. Each of these, in addition to upgrading, will also receive at least one new beam line and associated new facilities.

Superconducting magnets have already played a major role in the upgrading of the beam lines. The beam lines leading to the meson and proton areas have to be bent by dipole magnets to the left and right, respectively, of the straight beam line to the neutrino area. Fermilab's first large superconducting magnet experiment was the "left bend," a string of 22 magnets that began operating a year and a half ago under the direction of Roger Dixon of Fermilab. Among the lessons learned in commissioning the left bend was how important the helium refrigeration system can be. With the Energy Saver taking up most of researchers' attention, the left bend refrigeration system was originally barely adequate. Only after it was improved did the left bend work well.

These kinds of details are the sort of thing that Lederman refers to when he says Fermilab is "running scared" as the time to break in the Energy Saver approaches. While there is a feeling of hopeful expectation, Lederman says that the new superconducting synchrotron "has the highest risk of something going wrong of any accelerator so far."

—ARTHUR L. ROBINSON

Long-Awaited Decision on DNA Database

Molecular biologists can look forward to having access to a national DNA database now that NIH has at last awarded a $3-million contract

A long period of indecision and uncertainty in the field of molecular genetics has been brought to a close with the award of a $3-million, 5-year contract to Bolt, Beranek and Newman, a Cambridge-based company with expertise in computer communications. The company, which is subcontracting Los Alamos National Laboratory with Walter Goad as principal investigator, will be responsible for compiling and distributing a national database of DNA sequences.

The need for the database was first perceived urgently in 1979, but various circumstances combined to delay a decision on a national scale until now. Meanwhile, several groups around the country, and in Europe and Japan too, began independently to establish rival facilities. Although there has been considerable cooperation between the Europeans and groups here, their main effort, at the European Molecular Biology Laboratory (EMBL) in Heidelberg, came to fruition in April this year when their Nucleotide Sequence Data Library became freely available. EMBL's early success causes considerable embarrassment on this side of the Atlantic.

Ever since DNA sequencing became virtually a routine procedure in most molecular biology laboratories, the need for handling the inexorable information tide became obvious and pressing. Practitioners needed a way of compiling sequences with some kind of standard annotation showing known functional regions, such as transcription signals and splice sites. At least as important, however, is the facility to search for homologies between a new sequence and all existing sequences and to do sophisticated analytical procedures that might reveal other significant regions.

Until the middle of last year, when budget tightening intervened, the National Institutes of Health (NIH) had intended to support compilation and distribution of the DNA database as one project and the development of analytical software systems as a second, related, project. The contract to Bolt, Beranek and Newman covers just the first of these two, with the second, which might have cost around $2 million over a period of 5 years, on hold, perhaps indefinitely. As a result there seems likely to develop some sharp competition in originating and offering analytical expertise.

The momentum to establish a national DNA database got under way in March 1979 when the Rockefeller University was host to a workshop sponsored by the National Science Foundation. Carl Anderson of the Brookhaven National Laboratory had initiated the meeting and was its chairman. He recalls a unanimity that something should be done, but a diversity of opinion as to how to proceed. Should there be an extensive computer network, with centralized data collection and storage together with a big effort for the development of analytical software? Or would a much more modest and limited data-collection system be more prudent, at least to start with?

Opinions were sharply divided between these two conceptions. In addition, many people were uncomfortable with the prospect that sequences might become freely available before principal investigators had had time to work with them and therefore benefit from their sequencing efforts. Concern over potential violation of a researcher's right to prior access to his or her own data was deeply felt. Computer anxiety was still strong in the molecular biology community at the time.

For these several reasons the impetus of the Rockefeller meeting was dissipated, thereby exacerbating rather than avoiding one of its prime aims—avoiding the duplication of data-collection efforts.

By the end of 1979 the NIH had received proposals for data collection and analysis from several groups, including those of Margaret Dayhoff of the Biomedical Research Institute, Washington, D.C., and Walter Goad at Los Alamos. Dayhoff and her group already had considerable experience in compiling and distributing data on amino acid sequences of proteins. The Los Alamos group had begun serious work on DNA sequence collection in mid-1979.

As these several proposals involved a substantial service component in addition to research, the NIH realized that its support for any national facility would have to be by contract rather than as a research grant award. Primarily at the instigation of Elvin Kabat of the National Institute of Arthritis, Diabetes, and Digestive and Kidney Diseases, a series of NIH workshops was organized by the National Institute for General Medical Sciences (NIGMS). The aim was to prepare the ground for two requests for proposal (RFP's) for a national DNA data facility and for sequence manipulation software. In the meantime, Dayhoff and Goad received separate small grants to develop a system for data collection as an interim measure.

The first workshop was held at NIH in mid-July 1980, 1 month after the EMBL had decided to set up its own sequence data center. Subsequent workshops, in

0036-8075/82/0827-0817$01.00/0

October and December, occurred more or less in parallel with meetings in Heidelberg so as to promote collaboration across the Atlantic. Gregg Hamm, who is responsible for the EMBL's library, attended the October and December workshops at NIH.

The original notion was that the European and U.S. systems would be compatible in compilation characteristics and that data collection would be divided equally between the two centers. Information would then be exchanged so that both centers would maintain complete databases. Because of tardiness on the American side, the EMBL decided to go it alone, for the time being at any rate. Eventual division of effort is still a target for both sides.

The causes of delay were several. First, although the project was being shepherded along by the NIGMS, it was clear that other agencies would have to contribute financially. Bureaucracy being what it is, the necessary meetings between potential contributors took much longer than the molecular biology community would have liked.

One potential contributor, the Department of Defense, caused some practitioners a degree of nervousness. Why was the DOD interested? Would a military flavor insinuate itself into this project? Elliot Levinthal of the Advanced Research Projects Agency explains that the DOD has a continuing interest in possible military applications of recombinant DNA technology, in developing new materials or chemical sensors for instance, but not, he insists, in chemical or biological warfare. But when the NIH decided it would not for the present support development of analytical software for the DNA database, the Defense Department's enthusiasm for the project cooled. The department's cooperation is, however, required in allowing Los Alamos and Bolt, Beranek and Newman to use ARPANET, a nationwide-and-beyond network of rapid computer links, to communicate with each other.

In addition to time-consuming entanglement in multiagency bureaucracy, the change of administration at the beginning of 1981 conspired to bring further progress almost to a complete halt until uncertainties over budgets and the fate of certain agencies were settled. Two RFP's were produced, one for data collection and distribution, the other for data analysis. Eventually only the first one emerged, on 1 December 1981. The second fell casualty to restricted budgets and to the hope that demand for the expertise might be fulfilled from other sources.

The prime contenders for the contract were Dayhoff's group, Bolt, Beranek and Newman, and Intellegenetics, a newly formed Palo Alto–based company specializing in the application of data processing and artificial intelligence techniques to biological problems. Both Bolt, Beranek and Newman and Intellegenetics offered themselves as experts in data distribution, with the Los Alamos group providing the data-collection facility.

Even though Intelligenetics was a newly formed company, it was thought to be a strong candidate because of its origins in Stanford research of personnel and its concentration on computer applications to molecular biology. There was considerable surprise, therefore, when the company was eliminated at a preliminary technical evaluation stage. Only Dayhoff and Bolt, Beranek and Newman went through to final selection. The Cambridge company's long-established expertise in computer communications was probably influential in its successful bid for the contract.

Computing is fast becoming an integral part of molecular biology

Although not as deeply committed to molecular biology as Intelligenetics, Bolt, Beranek and Newman have been involved in biomedical computing since the early 1970's, when it set up the NIH-supported PROPHET system. User's access to the DNA database will be independent of PROPHET, but the PROPHET network will be the basis of a convenient distribution system.

The Los Alamos database currently contains some 600,000 sequences, which is about two-thirds of the total available. Goad says that the library will be fully up to date within a year and thereafter will collect new sequences within 3 months of their publication. Until recently new sequences were accumulating at an accelerating rate, but that has tailed off somewhat and is now running at an annual rate of about 400,000. Acceleration can be expected to return soon, however, especially with the anticipated publication of some large sequences, such as the 40,000 bases of the entire lambda genome from Fred Sanger's laboratory in Cambridge, England. The sequence of Epstein-Barr virus, with 170,000 bases, will follow in about a year.

Users will have access to the data online, by magnetic tape, or in hard copy, for which initially at least there will be a modest user fee. If the facility proves useful, says one NIH official, academic investigators might be willing to pay more for it after a year or so, thus covering more of the cost. Industrial users will pay higher fees from the start.

Valuable though the database might be in itself, the ability to perform sophisticated sequence manipulation would make it even more so, as evidenced by the recent experience on the NIH-supported Stanford University Medical Experimentation Computer Facility, SUMEX. A number of Stanford researchers, including Lawrence Kedes and Edward Feigenbaum, developed some artificial intelligence–based software for molecular biology analysis on SUMEX. Known as GENET, this package was opened as a guest account on SUMEX in March 1980. Commercial biotechnology companies flocked to use it, so much so that by September the same year the load was so great that a complete halt on commercial use had to be called. This had two consequences. First, academic use very quickly expanded. Second, Kedes, Feigenbaum, and two Stanford colleagues established Intelligenetics, which lincenses GENET programs for commercial use on its own computer.

Academic use of the GENET account on SUMEX has now reached the point where its volume too is a problem, and an NIH-appointed panel is soon to announce severe restrictions. Customers will be advised to seek alternatives, such as the facilities offered by Intelligenetics. Meanwhile, Bolt, Beranek and Newman, Goad, and another group headed by Michael Waterman, until recently of the University of California, San Francisco, are each intent on developing software for the type of analysis that was envisaged by NIH in its now defunct second project. Waterman has recently been awarded $800,000 over 3 years from the Systems Development Foundation, Palo Alto, to pursue this. Competition is likely to be keen.

Computing is fast becoming an integral part of molecular biology, and the trend is certain to continue as the flood of sequence data swells, as the need for extracting meaningful information from it grows, and as computer hardware becomes ever cheaper and more powerful. The arrival of the computer era is emphasized by an NIH-funded 3-week workshop on Computers in Molecular Genetics to be held in Aspen, Colorado, in September. Computer anxiety is surely past.—ROGER LEWIN