

## Human genome

## Department of Energy on the map

Washington

UNTIL recently, sequencing the human genome seemed an impossibly large project, almost like counting all the grains of sand on all the world's beaches. But today the Department of Energy (DoE) is giving serious consideration to a project to sequence all of the estimated 3,500 million base pairs making up the human genetic material. Initial estimates put the cost of the entire project at \$1,800–\$3,600, or about \$1 per base pair.

Early in March, DoE held a conference in Santa Fe, New Mexico, to discuss the feasibility of undertaking the sequencing project. The conference, hosted by DoE's Los Alamos National Laboratory, produced a report outlining a strategy for accomplishing the task. According to David Smith of the Office of Health and Environment Research (OHER) at DoE, the department is expected to decide in a matter of weeks how and whether to proceed with the monumental project.

DoE has historically had an interest in the human genome, and has been supporting research into genetic damage from energy sources for some time. Together with Lawrence Livermore National Laboratory, Los Alamos has been involved in the National Laboratory Gene Library Project, an effort to construct a chromosome-specific gene library. So far, the small insert library — containing fragments of 5,000–10,000 base pairs — is complete, and work on the large insert library for fragments of up to 50,000 base pairs has begun. Los Alamos has also developed flow cytometry technology capable of rapidly separating large amounts of DNA by chromosome, a capability that Los Alamos senior fellow Mark Bitensky sees as important to the success of the project.

Bitensky is coordinating Los Alamos' effort to make the gene sequence a reality. He says the project will "require many laboratories, many funding agencies, and many technologies", but he believes it can be done, and that the time is right to do it.

Although DoE may seem like an unlikely agency to handle such a project, Bitensky and OHER head Charles DeLisi feel that DoE is well placed to take a lead role. In addition to the gene library project, DoE has the computational facilities necessary for such a large-scale project. And DoE believes it has the know-how to undertake large-scale scientific projects, pointing to the DoE-run Fermi National Laboratory as an example of its expertise.

Current plans call for a decentralized programme, with one coordinating facility organizing the work of many centres around the country. The organizers hope the sequence can be a multinational

effort. Bitensky says European interest in the project is already high.

Two theoretical approaches to the sequencing project were considered at the Santa Fe meeting. One involved randomly sequencing fragments from the genome, and trying to determine the order of the fragments on a *post hoc* basis (see below). This was rejected in favour of a scheme where the genome is first sorted by chromosome, then digested with a restriction enzyme like *Sfi* I that will produce a relatively small number of relatively large fragments. The first task would be to order these large fragments by chromosome, with a small, pilot sequencing effort started while the ordering task was being completed. Having ordered fragments not only breaks the sequencing down to a size more easily managed conceptually, but will also provide a useful tool for investigating genetic diseases, Bitensky believes.

Critics of the sequencing project do not deny that the information would be useful. Rather they worry that spending so much money on such a large project may be an inefficient use of funds. George Church of the University of California at San Francisco, a participant at the conference, echoed the widely held concern that such a large project could draw money away from other research projects. Church is working on new DNA sequencing techniques, and would like to see the project proceed. Others, including Fred Blattner of the University of Wisconsin, feel that a complete sequence from a smaller organism would yield a great deal of information about cellular processes at a small fraction of the cost. Blattner is currently working on sequencing the *Escherichia coli* genome, and others are working on the yeast and nematode genomes. Another option would be to sequence just part of one human chromosome, providing a window into human genetic control without devoting so many resources to such a major project.

Still unresolved is what to use as a starting material for the sequence project. One possibility is the DNA from a hydatiform mole. Such DNA can be produced in large quantities, but may be inappropriate as it has aberrant genetic properties. A second possibility would be to use material from a variety of different sources, perhaps a different individual for each chromosome. But critics worry that any choice will necessarily exclude information on the diversity of human genetic material.

Bitensky says that one possible method for selecting the DNA to use mentioned at the Santa Fe meeting was to hold a competition among the world's men and women. The person coming up with the largest contribution to the project would have his

## The numbers game

Washington

SEQUENCING the estimated 3,500 million base pairs in the human genome is a mammoth undertaking. But Fred Blattner of the University of Wisconsin came away from the Department of Energy's March meeting on the sequence project optimistic about the prospects for completing the task in a reasonable length of time.

The approach that encourages Blattner does not require ordered clones or restriction maps. Instead, sequencing starts on random pieces of DNA. Sorting out repeating sequences may not be possible using this approach, but all the unique sequences would be uniquely assembled.

To show this approach is possible, Blattner presents the following scenario: Work-stations would be set up around the country to carry out the project. Each station would consist of two technicians and four automated sequencing machines. The machines would have 40 lanes, with each lane capable of recording 750 residues. Running the work stations in two eight-hour shifts five days per week gives:  $750 \text{ residues} \times 40 \text{ lanes} \times 4 \text{ machines} \times 2 \text{ shifts} \times 5 \text{ days} = 1.2 \text{ million base pairs per week, or } 60 \text{ million per year.}$

To make sure the entire genome is covered, Blattner figures it will be necessary to calculate 10 times the number of residues in it, or 35,000 million base pairs. Using this scenario, 580 work-stations could complete the job in a year. With each station costing \$500,000, the project would cost \$290 million.

Automated sequencers with such capabilities will soon be on the market, says Blattner. Although the information produced by this random sequencing approach would not be as rich as a complete ordered sequence, it would still be extremely useful. A later step could be to put all the unique segments in order.

Despite the organizational problem of coordinating the work of so many laboratories, Blattner believes it can be done.

Joseph Palca

or her DNA sequenced.

A major player in the sequencing project could be the Howard Hughes Medical Institute, which is already supporting restriction mapping work by Raymond White at the University of Utah Medical School and Frank Ruddle at Yale University. The institute's president Donald Fredrickson confirms that there is interest in the DoE project, but wonders if the time is right to begin such an endeavour.

DeLisi, described as the driving force behind DoE's efforts on this project, says nobody gave serious thought to the idea of sequencing the entire genome as little as five years ago. Now, it is technically feasible: all that is required is the will and the money.

Joseph Palca