New Game Plan for Genome Mapping

A new proposal, to be aired next week at the Human Genome I meeting in San Diego, promises to transform efforts to map the human chromosomes

"I WAS SKEPTICAL that we could create a plan that would not be laughed at," recalls Norton Zinder of Rockefeller University. "The whole community is out there waiting to nitpick and backbite. . . . [But] when I walked away from that meeting I knew we had it. I felt good."

Zinder had every reason to be elated. A proposal, presented to a select group of biologists at Cold Spring Harbor Laboratory in late August, could transform efforts to map the human genome (*Science*, 8 September, p. 1036). The proposal, in essence, provides a way to bring together the results of an array of different mapping techniques that had seemed incompatible. As a result, the initial goal of the human genome project suddenly seems both clearer and attainable.

The idea, outlined on page 1434 of this issue, will get its first public airing at next week's Human Genome I meeting in San Diego.

The initial goal of the genome project is to develop a physical map of the human chromosomes within 5 years. The problem is that "no one has defined in a technically credible way what the physical map of the human genome that we are supposedly constructing will look like," says Maynard Olson of Washington University, who is one of the nation's premier mappers.

True, the broad outlines are clear—a physical map shows the actual distance, ideally measured in nucleotide bases, between landmarks distributed along the chromosomes. Genes can then be located within those landmarks. But researchers constructing pieces of this map have yet to agree on what the landmarks should be. And without a common set of landmarks, mapping the chromosomes is a bit like building a road through a mountain: if tunnelers at both sides don't use the standard benchmarks that mark elevation from sea level, they're likely to end up with shafts that don't meet.

The new proposal, put together by Olson and three of his colleagues—Leroy Hood of Caltech, Charles Cantor of Lawrence Berkeley Laboratory, and David Botstein of Genentech—is a fresh approach to physical mapping that not only provides a clear definition of what the physical map should look like but also offers a common language and common set of landmarks.

"Everything written [about physical mapping] is already obsolete," exults Zinder, who chairs the NIH Program Advisory Committee on the Human Genome. "It has changed the whole outlook on the problem."

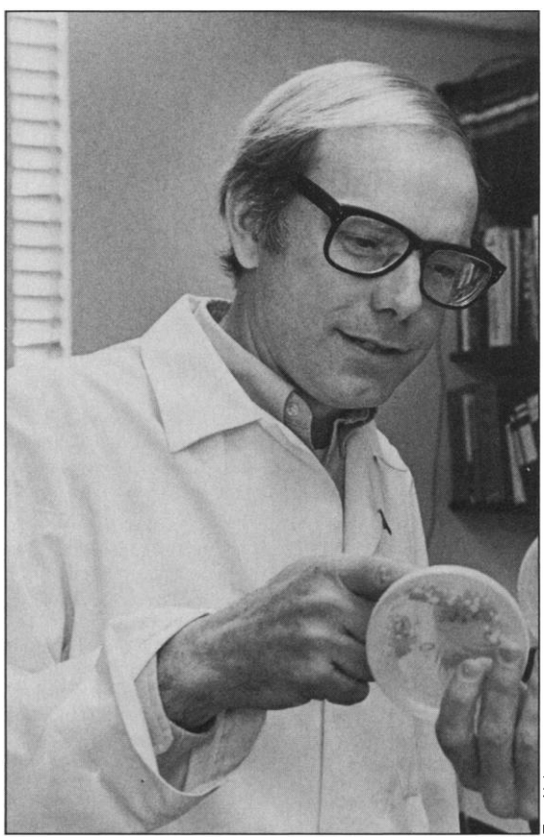
The idea is simply to use short, tagged tracts of DNA sequence as the landmarks in the physical map. As Olson explains, this is not a new mapping technique, though it is intimately tied to the new technique called polymerase chain reaction, or PCR. Rather, it is a policy proposal—a plea for mappers to record their results in the same language, no matter what techniques they are using. And with this new approach, a surprising number of problems disappear.

Indeed, if the enthusiastic claims are borne out, this new approach—dubbed STS for sequence tagged sites—will not only make it possible to integrate data from various mapping methods but will facilitate cooperation and sharing among labs. It will do away with the need to exchange clones—copies of pieces of DNA that are the currency of physical mapping—or to store them into perpetuity. And it ensures a place for "little" science in the mapping project, which may help to defuse tensions between NIH and the Department of Energy, which are still quibbling over the merits of the "big" and "little" approaches.

The scheme met with rave reviews when Olson presented it at Cold Spring Harbor, where 25 prominent biologists convened to plan the future, or at least the next 5 years, of the human genome project. "It carried the day," says Zinder, who notes that there were no skeptics among the assembled biologists, who are not known for agreeing on anything.

If it is greeted with the same enthusiasm at next week's meeting, the proposal seems likely to become the centerpiece of the 5-year plan NIH and DOE must present to Congress this February. The four authors of the proposal are certainly in a good position to see that their scheme is quickly translated into policy, for they serve on either or both the NIH and DOE advisory committees for the genome project. They also wrote the physical mapping section of the influential National Research Council report on the genome project nearly 2 years ago.

What has hampered progress toward a complete physical map to date is that there is not one type of physical map but several, and mappers are concocting new strategies at a fair clip. All of them involve cutting the DNA into pieces and then reordering the pieces as they would appear along the chromosomes. Mapping thus involves determining where a probe or piece of DNA fits on a chromosome. Once complete, a physical map will enable investigators to pinpoint a gene of interest, say, a



Tom Heine

Prime mover. *Gene mapper Maynard Olson has so far received rave reviews for the proposal.*

disease gene, to a particular fragment and, eventually, to pull it out and sequence it—that is, determine the exact order of its nucleotide bases.

The problem is that each type of map has its own language, if you will, and its own landmarks, which makes integration a nightmare. In restriction maps, the landmarks are the sites where a restriction enzyme snips the DNA. In "contig" maps, the landmarks are the overlapping ends of each clone, and so on. Thus, no matter how good each map is, the various kinds cannot be readily combined to create a larger, complete map of the human genome.

This situation has spawned numerous arguments on which mapping method is best, says Olson—arguments that he believes miss the point. "I don't want a consensus on the means," says Olson. "I want the strongest labs to do what they do best. The idea of letting 1000 flowers bloom is fine for the means. But this eclecticism has been allowed to spill over into the goal, and that has the potential to spell disaster."

The beauty of this new approach, the four authors say, is that it allows labs to use whichever mapping techniques they choose as long as they convert their results into a common language. It also provides a clear definition of what the map should look like.

They are proposing, as the new 5-year goal, a 100-kilobase STS map—that is, a map with these new landmarks spaced roughly every 100,000 bases apart along all the chromosomes. Given that there are some 3 billion bases in the genome, that means that 30,000 of these sites will need to be defined and mapped to the chromosomes.

Says Olson: "I was reluctant to support a 5-year goal because I wasn't sure what we were trying to do. Now it starts to make sense to talk about 5 years."

The gist of this proposal is that an investigator must simply agree to work out the nucleotide sequence of a little bit of the piece of DNA he has mapped. That short sequence, the sequence tagged site, then becomes the landmark on the map. And once the sequence is recorded in a database, any researcher can quickly recover—recreate, if you will—that piece of DNA without any biological materials ever changing hands. The logic of using sequence tracts as the landmarks seems unimpeachable, the authors say, because the ultimate physical map of the human genome is the exact sequence of all 3 billion nucleotide bases.

What makes this approach possible now, as opposed to several years ago, is a new technique called the polymerase chain reaction, or PCR—a means of amplifying, or making numerous copies of, DNA in a test tube. "We couldn't propose this 3 or 4 years ago; it wasn't technically credible," says Olson. "PCR transforms the project because it hands you the sample," adds Cantor.

The scheme would work this way. Once an investigator has mapped a piece of DNA, say a 40,000-base-pair clone, to a chromosome, he would then sequence a 200- to 500-base-pair stretch of that DNA, probably from the end. "With automated sequencing, that requirement is not so onerous," notes Cantor. The task is made easier still because the sequence need not be perfect; 98% accuracy is sufficient.

The next step would be to search at both ends of that 500-base-pair tract for two short, unique sequences, each about 20 nucleotides long. Those two short sequences would then be synthesized—an automated task that is "push-button technology," as Olson describes it—to create the so-called PCR primers. These short primers can be used in a separate reaction to make copies of the entire 500-base-pair tract.

"Everything written [about physical mapping] is already obsolete. . . . It has changed the whole outlook."

—Norton Zinder

In a PCR reaction, the two primers are essentially added to a pot with a piece of DNA that serves as a template and with DNA polymerase, an enzyme that triggers DNA synthesis. The primers then seek out their complementary spots on the template and begin churning out numerous copies of the target DNA that lies between them.

In one fell swoop, this approach does away with the need to exchange clones—the bane of much of molecular genetics these days—and it makes the map data accessible to anyone, big and small lab alike. The sequences of the two primers, as well as that of the larger interval, would be reported to a database along with information about the location of the clone on the chromosome. That done, an investigator need only call up the database, synthesize the two primers, and recover the STS overnight. Indeed, says Olson, there is no need to even talk to the original investigator.

To James Watson, director of NIH's genome program, that is the biggest selling point of this strategy. "If you publish these [sequences], you don't have to wait 6 months to get the clone. I see it as a way of distributing information. We keep hearing complaints that 'I can't get a probe from lab X.' The researcher in lab X says he doesn't have a secretary to send it. It's a marvelous excuse."

This approach also circumvents the need for a massive, permanent repository for the clones used in physical mapping, an idea that Hood and Olson characterize, respectively, as "crazy" and "dumb." The assumption has been that, when full-scale sequencing of the human genome begins in 10 or 15 years, investigators would use the stored clones as the starting material. But storing, cataloging, and distributing the 600,000 or so clones that might be needed could cost an estimated $30 million—more, in fact, than it would cost to create them. And that, Olson and his colleagues argue, would be a colossal waste of money, since no one would want to use 1980s clones in the 1990s anyway—and why should they, now that they can reisolate them with the aid of an STS.

"Here the repository is a computer with PCR primer sequence," says Hood.

And with the STS in hand, it is simple enough to get hold of the larger clone—say, the 40,000-base-pair clone—from which it was drawn. An investigator need only radioactively label the STS and use it as a probe to screen a "library," or collection of clones, to find that one piece.

The four are urging NIH to start now to "retrofit" existing landmarks by converting them into sequence tagged sites. Not all existing markers need to be converted, they say, only the 2000 or 3000 useful ones for which map positions are already well known. Olson estimates that the job would take about 3 years and cost roughly $10 million.

Converting these markers into tagged sites and then ordering them along the chromosomes would create a complete—albeit very low resolution—physical map of the human genome. "It won't be a very good map," concedes Olson, "but that is how the genetic map evolved. You start with something that is not too far wrong. It is hard to write on a blank board." The map would be filled in as data accumulate from labs around the world.

The other, unexpected upshot of this proposal is that it levels the playing field. Physical mapping had seemed stacked in favor of "big" science, says Olson, because the problem seemed so imposing that only big labs, like the national labs, would tackle it.

The big science vision of mapping calls for fragmenting DNA and then blanketing an entire chromosome with overlapping pieces—a mammoth undertaking. Two of the national labs are well under way on these contig maps for chromosomes 16 and 19.

In that scheme, however, there is little

room for the individual investigator who has mapped one small region of a chromosome in detail. In contrast, with the STS approach, data from any mapping endeavor, no matter how small, can be readily added to the evolving map. "It gives the individual investigator the power to map things," says Botstein. "He doesn't have to join up with Los Alamos."

Technically, there appear to be few obstacles. "It's a good strategy," says Henry Erlich of Cetus Corporation, one of the developers of PCR. "There are potential problems that can be imagined, but there are potential solutions too." Occasionally, Erlich says, the PCR assay is not perfectly specific; it amplifies DNA in other places in the genome that are similar to the two primers. But it is simple enough to build in "fail-safe" measures, he says, like adding an additional primer between the other two.

Ultimately, success will depend, of course, on people sequencing that bit of DNA—the 500 or so bases—and reporting it to the database. To Botstein the approach is self-implementing: "If people want to play in the arena, they will have to do this."

Olson, however, is less inclined to leave it to good intentions or peer pressure. Instead, he thinks that reporting map data in STS language should be a requirement of the genome project. "The genome project is trying to develop a physical map. It is reasonable to ask people to report in a common language."

DOE's support is crucial, as it is funding the biggest mapping efforts under way. So far, the reception has been enthusiastic.

"It's a terrific new concept," says Ben Barnhart, manager of DOE's genome program. "I certainly hope the scientific community adopts it. But it is not something you can impose."

"No question, we'll try it out in-house," says Robert Moyzis, who heads DOE's genome center at Los Alamos, where chromosome 16 is being mapped. "Los Alamos has the largest contig mapping project going, so we are in a good position to see how well this [STS] approach works." Moyzis, who was at the Cold Spring Harbor meeting where Olson first described the proposal, calls it a "conceptual breakthrough," though he predicts "there will be further iterations at further meetings. We need a way to be able to talk to one another and compare data. We need a mutual language, and this is likely to be it."

How far the proposal actually goes—and what, if anything will be required—will be hammered out in the hallways at San Diego and in the closed-door meetings as DOE and NIH plan their strategy for the next 5 years. ■ **LESLIE ROBERTS**

Conflict Over Conflict of Interest

If your spouse has ten shares in K Mart Corporation, should you be forced to disclose that fact the next time you file a grant application with the National Institutes of Health? If new draft guidelines on conflict of interest published earlier this month are adopted, you certainly will. And that's just for starters.

NIH, which developed the guidelines along with the Alcohol, Drug Abuse and Mental Health Administration, is tackling the conflict-of-interest question head-on because of what associate director for extramural affairs George Galasso describes as a climate that requires it to do so. And if NIH doesn't come up with strict requirements, Congress may step in with even stricter legislation. NIH signaled that it was taking the issue very seriously when it convened at a 2-day meeting on the topic on 27 and 28 June (*Science*, 7 July, p. 23).

The proposed guidelines would require anyone involved in NIH- or ADAMHA-funded research—as well as their spouses, dependent children, and other dependents—to make "full disclosure of all financial interest and outside professional activities" to their host institution. This information is to be provided by everyone receiving or applying for money from NIH or ADAMHA and is to be updated at least once a year. The guidelines would also prohibit anybody involved in an ADAMHA- or NIH-funded research project (or their dependents) from having "personal equity holdings or options in any company that would be affected by the outcome of the research or that produces a product or equipment being evaluated in the research project." Researchers would also be barred from receiving honoraria from companies whose products they are testing. Universities would be permitted to grant waivers from these restrictions, but the waivers would have to be reviewed by NIH.

Many worry that conflict-of-interest issues are too complex to resolve with such a sweeping but basically simplistic set of restrictions. "It is silly," says Carol Scheman of the Association of American Universities. "It is a misapprehension of what research is all about." Scheman argues that conflict of interest is essentially inescapable, and that what is needed is a more thoughtful set of principles that spells out which conflicts society will tolerate and which are unacceptable.

David Blake, associate dean for administration and planning at Johns Hopkins University School of Medicine, worries that the rules regarding consultantships will ultimately have a chilling effect on pharmaceutical companies that have come to rely on university researchers for advice. "We really need an economic impact statement on that one," he says.

Blake says universities will also have problems keeping track of the proposed disclosure of financial information. "It's a tremendous administrative burden for very little yield," he says.

Blake also believes that the mechanism NIH used to promulgate its proposals—it simply published them as guidelines in the 15 September issue of the NIH guide for grants and contracts—is an attempt to sidestep bureaucratic procedures that must be followed in issuing formal regulations. But Robert P. Charrow, formerly in the Department of Health and Human Services general counsel's office and now with the law firm Crowell and Moring, says guidelines that tell institutions what they shall and shall not do are regulations, like it or not. As such they must be published in the *Federal Register*, signed by the Secretary of Health and Human Services, and comply with the terms of the Paperwork Reduction Act, among other requirements. NIH has not followed any of these steps, and Charrow believes the guidelines would be nullified if anybody cared to mount a legal challenge.

Despite these problems, NIH has won some praise from Capitol Hill. Representative Ted Weiss (D–NY), whose hearings on conflict of interest focused attention on the issue, calls the proposals "an important step forward in dealing with this growing problem." But he says he is concerned about how NIH will punish institutions or individuals who violate the conflict standards, and he worries that universities may abuse their waiver rights for favored faculty.

"When the federal government is paying for the research, that research should not be tainted by any possibility of bias due to financial conflicts of interest," he says.

NIH has asked for comments on its proposals by 15 December. "Keep in touch," says Blake. "I'm sure this topic's going to be alive all year." ■ **JOSEPH PALCA**