Will "DNA Chip" Speed Genome Initiative?

New computerized methods of sequencing DNA promise great advantages but also tough technical challenges

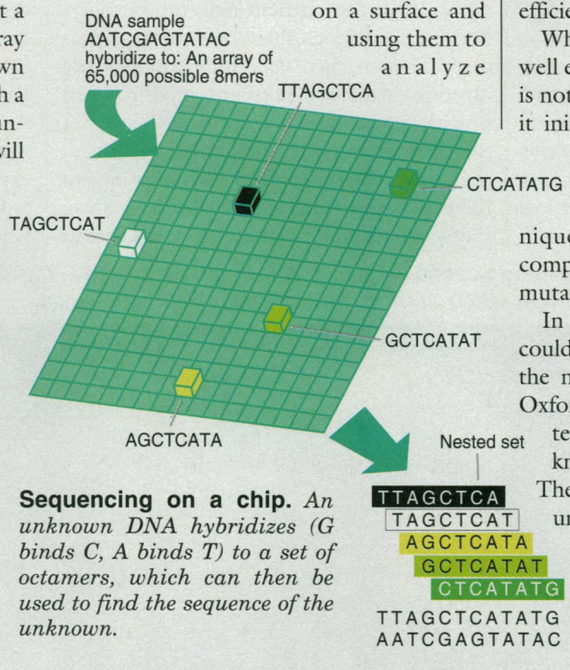
ONE OF THE BIGGEST CHALLENGES FACING researchers on the Human Genome Project is finding faster ways to sequence DNA. Current automated methods turn out about 2000 bases of finished sequence per day—a rate that would require 1.5 million person-days to sequence the whole genome. So the search continues for a means of boosting sequencing rates by orders of magnitude. "We need far-out approaches that could potentially be very fast," says Stanford genome researcher Ronald Davis. And they may have found one in the "DNA chip," a clever but as yet unproven idea with the potential, according to Charles Cantor, principal scientist in the Department of Energy human genome project, to "generate sequence at prodigious rates," 100 times or more faster than now possible.

Despite its nickname, the "chip" is not a piece of electronic equipment, but an array of short pieces of DNA—each with a known sequence—arranged on a substrate. Splash a solution containing a longer DNA of unknown sequence onto the array, and it will react with some of the short DNAs there, yielding a pattern from which a computer can deduce its sequence. At least that's the idea—most of the concept remains on the drawing board. A pilot experiment at Argonne National Laboratory used a variation of the method to deduce a 100-base sequence—a far cry from a whole genome or even the average gene (which has more than 1000 bases)—but enough of a start to tantalize biochemists and mathematicians alike. "A lot of good people have been attracted to the idea," says Cantor.

The underlying concept—independently conceived 3 years ago by several research teams—is disarmingly simple: If a solution containing a DNA fragment hundreds of bases long is exposed to all possible DNA pieces of a given length (say 8 bases), a subset of those "octamers" will be complementary to 8-base sequences somewhere in the DNA, and will bind (hybridize) to them. Those octamers make a "nested" set: the first is complementary to bases 1-8 on the unknown DNA, the second to bases 2-9, and so on. By using the overlaps to put the octamers in order, a computer could read off the full sequence of the unknown DNA.

There are 65,536 possible octamers, and that's a daunting number of hybridization reactions. But mathematicians are composing algorithms that may reduce the number of necessary octamers to thousands. And researchers are devising ways to do many reactions in parallel. Radoje Drmanac and Radomir Crkvenjakov, from the Institute of Molecular Genetics and Genetic Engineering in Belgrade, Yugoslavia, were among the first to articulate the idea. Now working at Argonne National Laboratory, they have designed an automated process for arraying 150,000 DNAs on filters that can then be exposed to a series of octamers or other DNA probes.

Then there's the "chip" approach, which relies on putting all the necessary octamers on a surface and using them to analyze different DNAs in series. Researchers at the Affymax Research Institute in Palo Alto, California, have devised a lithographic means of synthesizing peptides or nucleic acids on a glass surface. "We have just completed moving up to 65,000" sites in a 1.28-cm-square array, says Affymax scientist Stephen Fodor. Affymax developed the technique first for protein recognition, and is now working on DNA chips.



Sequencing on a chip. *An unknown DNA hybridizes (G binds C, A binds T) to a set of octamers, which can then be used to find the sequence of the unknown.*

If such a chip or one of its variations pans out, "it has the potential of being a very fast way to read sequences," says Ed Southern of Oxford University, who is also developing a DNA chip. "But we're a long way from that goal." Indeed, before such grand dreams can be realized, tall technical hurdles must be cleared. One of the most serious is how to handle an octamer that is repeated in the DNA. It can confuse the computer by acting as a "branch point" in the sequence being pieced together: Two different sequence paths lead into it, and two lead out, but there is no way to tell how those paths should be ordered. There is a 50% chance that a branch point will occur in a mere 256 bases of DNA, but mathematicians working on the problem believe they can solve it with clever octamer choice, and the addition of some longer probes.

Another serious obstacle stems from the hybridization reactions by which DNA binds to complementary sequences. The reactions can yield ambiguous results that would doom sequence analysis. Again, math may provide the solution, according to Robert Lipshutz, of the mathematical consulting firm of Daniel Wagner Associates, in Palo Alto. "It's a serious problem," he says, "but it can be treated as a statistical problem, and you can get very efficient algorithms to solve it."

Whether such problems can be solved well enough to allow large-scale sequencing is not yet clear. "I'm not that excited about it initially as a primary sequencing tool," says Stanford's Davis, who collaborates with Affymax. But even with its present flaws, Davis says the technique is well-suited to error-checking, comparative sequencing, and searching for mutations.

In all three of those cases, a DNA chip could be used for rapid comparisons without the need to decode entire sequences, says Oxford's Southern. You can predict the pattern that would be produced by the known sequence of, say, a certain gene. Then you can do the hybridization with an unknown—for example a mutant version of the gene. Subtract the two patterns, says Southern, and "what jumps out at you is the difference between them. It's very easy to interpret."

Though sequencing by hybridization shows promise, Cantor notes that few validating experiments have been done. In November, Cantor, Southern, and Andrei Mirzabekov, of the Engelhardt Institute in Moscow will host an "experts' workshop" in Moscow to clarify what needs to be tested. "This is an area that has had a lot of thinking and relatively few experiments," Cantor says. "Now we need some hard-nosed experimental validation. ■ MARCIA BARINAGA