



University of Oklahoma's Bruce Roe on Running Sequencers 'Without the Covers'

January 23, 2007



Bruce Roe
George Lynn Cross
research professor of
chemistry and
biochemistry, and
director
Advanced Center for
Genome Technology,
University of
Oklahoma

Name: Bruce Roe

Title: George Lynn Cross research professor of chemistry and biochemistry, and director of the Advanced Center for Genome Technology, University of Oklahoma (at Oklahoma since 1981)

Age: 65

Education: BA in chemistry, mathematics, and physics, Hope College, Mich., 1963
MA in chemistry and biochemistry, Western Michigan University, 1967
PhD in chemistry and biochemistry, Western Michigan University, 1970

Experience: NIH Postdoctoral Fellow, State University of New York at Stony Brook, 1970-1973
Assistant and Associated Professor, Kent State University, 1973-1981
Sabbatical Research Year, Frederick Sanger's Laboratory, Medical Research Council, Cambridge, England, 1978-79

As director of the University of Oklahoma's Advanced Center for Genome Technology, Bruce Roe oversees a group of nearly 60 people focused on an array of DNA sequencing projects. The center is currently sequencing four chromosomes of *Medicago truncatula*, a relative of alfalfa, as well as chimp chromosome 22, and microbial communities, among

other projects.

Roe's center also played a role in sequencing human chromosome 22 – the first human chromosome to be sequenced as part of the Human Genome Project. Roe himself has been involved in developing and popularizing Sanger's dideoxy sequencing method, and his team is still helping to push the frontiers of sequencing technology today.

In Sequence caught up with Roe recently to discuss the latest projects underway at his center, and his thoughts on current and future sequencing technology.

Tell me about the AdvancedCenter for Genome Technology. What is its mission, and what are some of the projects the center is currently working on?

We were originally funded as a genome center in 1990, one of the first three genome centers in the Human Genome Project. We sequenced genes involved in leukemia and then the first human chromosome, chromosome 22, in collaboration with the Sanger Center in England and Keio University in Japan. That was published in 1999.

Then we were part of the mouse genome project and sequenced the region in mice that's syntenic to chromosome 22. Right now, we are doing chimp chromosome 22, which used to be called chromosome 23 but is now renumbered to be correct. It turns out chromosome 22 is split into a and b in chimps, so they named them differently.

We are also sequencing plants. We are funded from the National Science Foundation as part of an international program to sequence *Medicago truncatula*, which is the laboratory cultivar of alfalfa. It's a diploid, but it's been inbred for so long, it's really monoplloid. We are doing four of the eight chromosomes here. It's close to 75 percent completed at this point in time and we are using a BAC-by-BAC approach rather than shotgun.

We are also in the legume community, sequencing several regions in soybean that are involved in disease-resistance. We are also trying to do some stuff now with cotton and some other plants, as well as looking at microbial communities. We have a whole group in the lab that's doing microbial community studies, everything from soil to oil

wells to water. In-depth studies, looking at not only the ribosomal RNA to see who is here but also looking at the proteome to see what genes are expressed, and doing EST studies in microbial communities.

Those are the bulk of the projects that we are doing scientifically. And along the way, we are also trying to do a lot of technique development.

What are the techniques your lab helped develop over the course of the last decade or so?

There is a history to that. When I went on my first sabbatical [in 1978], I went to the UK and worked in Fred Sanger's lab. I was part of the group that sequenced the human mitochondrial genome, and, along the way, helped develop the dideoxy sequencing method and make it fairly robust. Bart Barrell and Steve Anderson and several others worked on that project, and along the way, what I ended up doing was helping develop techniques for primer isolation and for really getting the sequencing moving along.

Then, when I came back to this country, there really had only been one person prior to me from the States in Sanger's lab doing something with the dideoxy method, and that was Clyde Hutchison [at the University of North Carolina at Chapel Hill]. When I came back, I was one of the original prophets in this country for the dideoxy sequencing method vs. Maxam and Gilbert's chemical approach. I had the second website that was set up at the University of Oklahoma, and we gave away the protocols from England because Fred Sanger wanted to give the protocols away for free.

Then Ellison Chen, my second graduate student, went to Bethesda Research Labs and developed the kit for dideoxy sequencing that was then sold commercially. Then he went to Genentech and Paul Armstrong, a postdoc of mine, went to take over from him at Bethesda Research Labs, and they marketed a kit for dideoxy sequencing. And eventually, the dideoxy method became very, very widespread in this country and eventually overtook the Maxam-Gilbert method.

I was always one of those guys who was freely available. Just like Allan Maxam was available to talk to people when they had questions about the Maxam-Gilbert method, I was always on the phone or e-mailing, answering questions about getting dideoxy to work. And then many other companies did things along the way, improved the enzymes, and eventually Applied Biosystems made this commercial instrument for DNA sequencing that came out of Lee Hood's lab. I was at a meeting at California and said, 'I'd really like to have that machine.' So the short of that story is that I gave them a purchase order and was the first site outside of Foster City to get the DNA sequencer 370 from ABI, in December of 1986.

Then we worked with Cheryl Heiner at ABI. They had a great instrument, but the protocols were not really up to date, and so we improved the protocols with Cheryl. And then again, we put those protocols out there for free. And that helped make the 370 much more of a robust instrument. And then we automated a lot of the upfront isolation of DNA, and automated the sequencing reactions, and still continued to stay at the forefront from the 370 to the 377. In fact, I wanted to run a longer gel, so I cut a hole in the roof of the machine so that we could stick a longer gel plate in, and made a box on top, covered it with black paper, so we could run meter-long gels on the 370.

We cannot treat science and the instruments that we have in science like black boxes. We have to go in there and open the cover and run the machine without the cover on to see what's happening, and really understand what's happening. I think too many people nowadays just run a machine, and they don't understand why it works or how it works. You ask the students a question, and they say, 'Oh, that's because I ran this program' or 'I used this machine and I got this answer.' Then I get rather upset with them and say, 'You have to know how the programs work and how the instrument and the machines work, and what's the theory behind them, and run them without the cover, and really get a feel for them, because that's the only way you can make them better.'

So over the years, I have been running things without the cover. We then went to the ABI 3700s and then recently the 3730s, and have been collecting a lot of data, and pushing the frontiers.

A year ago in December, we got one of the early releases of the 454 sequencer. And you know, immediately I took the cover off of that and said, 'This is silly; we can add more solutions and get longer reads,' so we are routinely now getting reads out past 300 bases, approaching 500 bases, very rapidly approaching 100 million bases in one afternoon's run. That's on the Genome Sequencer 20. We are scheduled to get an upgrade [to the GS FLX] this week, but the upgrade is basically going to do most of the stuff that we have already done, and [have] a few other small changes to the machine.

The real problem nowadays, with that instrument and with large-scale sequencing, is a software issue. We can get reads out to 500 bases with this instrument, and routinely get 300 bases, but the assembly software needs to be fine-tuned. As you get reads further out, you get slightly less signal, and then it becomes difficult to interpret, and so it becomes a software problem to figure out how to interpret the reads, or the flows, and how to weight them in the assembly. So we are working with 454 on some early releases of software to make it more robust and give them feedback. They are doing a very good job of trying to improve the software for these longer reads.

[We give our protocols for the 454 sequencer] away for free on the website, and contrary to ABI, who was upset when

we told the world how to dilute their reagents, 454 is really glad that I am putting this stuff on the website and helping people make the machines more useful.

New sequencing technologies are being developed as we speak. What is their greatest promise, and what are the challenges they still need to overcome?

I think the new sequencing machines are incredibly exciting. I think that the machines that give the very short reads — that are based on several different technologies, but you can get hundreds of millions of bases [per run], with very short reads of 20-some bases — are going to be really outstanding for SNP determination and for various resequencing projects and comparative genomics, where you don't really care about repeats and really care about unique regions. Those are going to find a place, I think, and be very, very useful.

Things like the 454, in my mind, and those machines that are coming out for longer reads, really are going to eventually replace the technology that we have with capillary machines. In fact now with the 454, we have paired-end reads, and we have long 454 reads, and we are really only using the capillary machines for finishing and PCR sequencing and stuff like that [here]. It's very expensive per sequence [base] to run the capillary machines, compared to the 454 machines. Though in an afternoon you spend roughly \$10,000 or thereabouts per run on the 454, and that's a lot of money, what we get is on the order of 70 or 80 million bases for that \$10,000.

These new machines that are now coming out I personally believe have a very small, finite window, because next-generation machines, like the improvements on 454-type pyrosequencing technology that are coming out of [Mostafa Ronaghi's group at] Stanford, or with the stuff out of Richard Mathies' lab [at Berkeley] and other labs — passing DNA as a single molecule through a detector to measure the sequence, and reading 20,000 bases or 100,000 bases, those kinds of really long, accurate reads — those are going to become [available] commercially, I would hope, within the next five years. That's going to mean that the present instruments are going to have this three- to five-year lifetime.

How are we going to sequence DNA 10 years from now?

You are going to hold one end, and I am going to hold the other end, and somebody is going to read it. We are going to be detecting fragments of DNA that are very, very long and get highly accurate sequence. The National Institutes of Health have done a really marvelous job with fostering [technologies for the] \$10,000 genome and the \$1,000 genome. People are using all kinds of nanotechnology and other kinds of really, really bleeding-edge forefront techniques to try and tackle this problem — it's incredibly exciting, and it's beyond for you and me, but when children are born 10 years from now, for \$1,000 we will hopefully be able to give them an idea what their genome looks like. The problem is getting insurance companies to pay for it, but that's sort of irrelevant, because the gain from that kind of knowledge, what kind of genetic diseases we are going to be prone to, and how to then prevent them, is very, very important. And that's going to really help and be just a boon to all of us.

© Copyright 2007 GenomeWeb Daily News. All rights Reserved.