2 chasing 1

By Lev Osherovich, Senior Writer

Since **Helicos BioSciences Corp.** launched its single-molecule sequencing platform in 2008, companies have been competing to improve the DNA sequencing technique's sensitivity, speed, read length and price. **Life Technologies Corp.** and **Pacific Biosciences Inc.** say they have achieved these goals by optimizing detection and reagent use. Pacific is already shipping its machines to early customers, and Life Technologies will make its first shipments by year end.

In single-molecule sequencing, also known as sequencing by synthesis, a DNA polymerase incorporates fluorescently labeled nucleotides along an individual, unamplified DNA template. Each of the four possible nucleotides carries a different fluorescent dye. The fluorescent emission pattern generated during this iterative process, in which the incorporated nucleotides are identified by the color of the emitted light, corresponds to the template's nucleotide sequence.

In contrast, conventional ensemble methods involve simultaneous sequencing using DNA probes or truncated synthesis reactions of many identical templates amplified from an initial sample.

"Single-molecule sequencing means that at the moment of detection, I am looking at one fluorescent molecule," said J. William Efcavitch, SVP and CTO of Helicos. "This is opposed to ensemble methods, where you're looking at hundreds of thousands or millions of molecules at a time."

The three companies' approaches all employ the same sequencing-by-synthesis principle but use very different detection schemes to collect the resulting data.

In contrast to Helicos' platform, in which the sequential incorporation of labeled nucleotides to the nascent DNA strand requires repeated rounds of adding, manipulating and removing reagents, the latest technologies avoid these steps by making the reaction and detection steps more self-contained. And, in a crucial advance, the DNA polymerase now not only attaches nucleotides to the new DNA strand but also triggers the release of a just-incorporated dye to allow the detection of the following nucleotide and its dye without having to manipulate or remove any reagents.

Pacific Biosciences announced the shipment of its first sequencing machines at last month's Advances in Genome Biology and Technology conference in Marco Island, Fla.

At the same conference, Life Technologies unveiled its own take on single-molecule sequencing. The company thinks its technology will complement its existing SOLiD ensemble sequencing system.

FRETs and dots

Life Technologies' approach involves real-time, continuous monitoring of DNA polymerase activity as the enzyme moves along a DNA template immobilized on a glass slide.

CTO Joseph Beechem told *SciBX* the company has designed a light-focusing quantum dot nanocrystal tethered to the polymerase (see **Figure 1, "Three methods for single-molecule sequencing"**). Together, the polymerase and nanocrystal form a miniature sequencing machine with a built-in signal amplifier.

The quantum dot absorbs light from a laser, emitting a second burst of light that excites a fluorescently labeled nucleotide. This transfer of energy, known as fluorescence resonance energy transfer (FRET), can only occur across a distance of about 10 nm, ensuring that only the nearest nucleotide—the one in the enzyme's active site—is excited.

The polymerase then moves along the template, expelling the first fluorescent dye and binding a new labeled nucleotide corresponding to the next position on the DNA. Reaction conditions allow recurring cycles of synthesis and detection to occur without external manipulation.

"The critical item here is building the sequencer on the surface of a quantum dot nanocrystal," said Beechem. "The signal arises from this portable reagent."

Beechem said the nanocrystal-bound polymerase can move along a DNA template of several thousand base pairs, reading out its sequence as a series of light flashes at a speed of several per second. Because the template DNA is tethered to a solid surface, a polymerase molecule that wears out after many repeated cycles can be replaced with a fresh one at any time to obtain yet longer reads, he added.

The company did not disclose what the maximum read length would be.

"We can get a turnaround of 30 minutes to an hour. This is a lot quicker than the 24 hours it takes to do the sample preparation and analysis" using ensemble methods or competing single-molecule approaches, said Andy Watson, VP of business development at Life Technologies.

However, Watson said the Life Technologies system will require specialized detection instruments that the company has not yet built for the general market. He said the machines should be available by the end of the year and predicted that the ultimate cost per experiment, encompassing one read-through of tens of thousands of DNA templates, would be "one to a couple of hundred dollars."

A well of information

The Pacific Biosciences approach involves spatially confining the template-polymerase complex to ensure a unique fluorescent signal by physically shutting out extraneous nucleotides.

In contrast to Life Technologies' tactic of attaching the DNA template to a particular spot on a slide, Pacific tethers an individual poly-

"We are seeing the fragmentation of the DNA sequencing industry."

**—Radoje Drmanac,
Complete Genomics Inc.**

Figure 1. Three methods for single-molecule sequencing. Researchers at **Life Technologies Corp.** have presented a way to rapidly sequence single molecules of DNA using fluorescence resonance energy transfer (FRET) from quantum dot nanocrystals. The method uses a DNA template tethered to a slide and a free DNA polymerase affixed to a quantum dot nanocrystal (I[a]). When excited by a laser, the nanocrystal emits a burst of FRET light (I[b]) that is absorbed by a dye-labeled nucleotide in the polymerase's active site. The dye then emits its own burst of light, which is recorded (I[c]). The polymerase subsequently cleaves off the dye and moves on to incorporate the next nucleotide (I[d]).

Pacific Biosciences Inc. and **Helicos BioSciences Corp.** also have platforms for sequencing single molecules of DNA.

Pacific Biosciences uses a DNA polymerase molecule tethered to the bottom of a nanowell (II[a]) that ensures only one nucleotide-linked dye can be directly excited at a time. The dye is removed by the polymerase when the next labeled nucleotide is incorporated (II[b]).

Helicos' method uses DNA templates tethered to a glass slide. The templates are extended with DNA polymerase and a single type of dye-labeled nucleotide (III[a]) that labels individual spots on the slide (III[b]). The slide is washed and photographically scanned to reveal where the dye was incorporated. The dye is then chemically removed (III[c]) and another dye-labeled nucleotide is added with fresh DNA polymerase (III[d]). Helicos' technology is already on the market.

merase to the bottom of a nanoscale well in a metal and silicon chip, and the data are captured as the DNA passes through the polymerase (see Figure 1, "Three methods for single-molecule sequencing").

The company's instruments detect fluorescent emission at the bottom of the nanowell, achieving single-molecule detection without the need for quantum dots or FRET, said founder and CTO Stephen Turner.

"We can also sequence with a tethered DNA molecule and a free polymerase, but this limits the read length because the polymerase starts to wander" out of the zone of detection, said Turner.

According to Turner, tethering the polymerase avoids the problem of

tracking the enzyme as it moves further along the template, which gyrates in ever-widening circles as the strand grows.

Improved read length "is the clear advantage of working with a fixed polymerase," he said, noting that Pacific presented reads of more than 10,000 base pairs at the Florida conference.

Turner also said use of a single, tethered polymerase leads to lower reagent costs compared with ensemble sequencing and the Helicos method.

During 1H10, Pacific Biosciences plans to deliver its first generation of sequencers to 9 academic centers and 1 private customer—

Monsanto Co.—at a cost of \$695,000 per instrument. Turner said it initially will cost \$99 to sequence 80,000 templates to a length of 5,000–10,000 base pairs and that the total sample preparation and run time is approximately 8 hours.

At this rate, the cost of sequencing 1 billion base pairs would be \$125–\$250—a human genome has approximately 3 billion base pairs and is typically sequenced 12 or more rounds to provide a high-fidelity sequence read.

Competitive advantage

Although the Life Technologies and Pacific Biosciences approaches are fairly similar in principle, they involve different technical features that could determine how the products are adopted in the marketplace.

“Pacific’s work depends on those nanowells, while Life has their DNA in an ordered array” on a chip surface, noted Helicos’ Efcavitch.

According to Efcavitch, the biggest challenge for both competitors will be to minimize the noise caused by the detection of unincorporated nucleotides in solution. “You need to distinguish the molecules that are in solution from the ones that are incorporated into the growing DNA strand. Both techniques are clever but it remains to be seen which has the lower error rate,” he said.

Helicos’ single-molecule sequencing platform includes washing steps that remove any unincorporated, labeled nucleotides before every detection step. The company’s instrument determines the DNA sequence at each position on the slide by analyzing serial images of the reaction.

Helicos’ instrument, which also can be used to sequence RNA, sells for \$799,000 and costs \$18,000 per sequencing run to operate. One sequencing run yields approximately 21–35 billion base pairs’ worth of sequence data, which, according to Efcavitch, translates into 3–4 runs per genome to ensure the highest accuracy of the final DNA sequence. Each run takes seven and half days, he added.

Clifford Reid, president and CEO of sequencing service company **Complete Genomics Inc.**, said that single-molecule approaches offer “high quality, low cost sequence,” but ensemble sequencing has evolved to remain cost-competitive with the new methods.

Complete Genomics uses conventional ensemble chemistry to sequence extremely short strands of DNA and computationally stitches them together to obtain continuous sequence data.

According to Reid, for whole human genome sequencing, Complete Genomics’ automated sample preparation system and economy of scale will trump single-molecule approaches when it comes to cost. The company’s platform is optimized for human DNA analysis and the company’s services therefore are focused on human sequencing. The company charges an average of \$10,000 to sequence an entire human genome, with lower prices available for bulk orders.

Life Technologies’ Watson also said single-molecule sequencing doesn’t scale up as effectively in this space as ensemble sequencing methods such as the company’s SOLiD technology.

With single-molecule sequencing, “we’re looking at tens of thousands of reads compared with the billion possible with SOLiD,” he said. For this reason, he expects ensemble methods to remain the

technology of choice for systematic genomic sequencing studies such as the recently reported sequences of individual patients with Charcot-Marie-Tooth disease.¹

Life Technologies spokesperson Patty Zamora said that it costs approximately \$6,000 to sequence a human genome using the company’s currently available SOLiD platform.

Life Technologies sees a role for the single-molecule technology in clinical diagnostics of repetitive DNA in medically relevant regions of the human genome. One example, said Beechem, is the major histocompatibility complex (MHC) locus, which must be checked for compatibility before transplant surgery. Ensemble sequencing methods have a hard time getting through the MHC region because of its highly repetitive sequence, a problem obviated by a long, continuous single read, according to Beechem.

“A lot of key medical decisions require this sort of deep sequencing,” he said. “This will hopefully be a machine that translates into personal genomics.”

Many winners?

Turner expects that the Pacific Biosciences technology will be cost-competitive for every type of sequencing, including that of whole human genomes.

“We will be able to address all of the sequencing markets,” he said.

However, Efcavitch and Complete Genomics CTO Radoje Drmanac both noted that the evolution of new sequencing technologies has led to the emergence of niche markets with specific technical requirements.

“We are seeing the fragmentation of the DNA sequencing industry,” said Drmanac.

In the near term, he said, the market has room for a range of devices such as “same-day sequencing boxes for \$50,000” as well as more expensive custom-configurable high throughput instruments and more specialized apparatus for clinical diagnostic sequencing.

“These are really different markets,” said Efcavitch. “People will use different platforms for different answers.”

He said the long reads made possible by single-molecule approaches may have applications in human genome diagnostics and certain areas of microbial research.

“Because single-molecule sequencing has a short turnaround time, infectious disease diagnosis will be another good area,” said Drmanac.

All three companies have patented their technology.

Oshrovich, L. *SciBX* 3(11); doi:10.1038/scibx.2010.331
Published online March 18, 2010

REFERENCES

1. Lupski, J.R. et al. *N. Engl. J. Med.*; published online March 10, 2010; doi:10.1056/NEJMoa0908094

COMPANIES AND INSTITUTIONS MENTIONED

Complete Genomics Inc., Mountain View, Calif.
Helicos BioSciences Corp. (NASDAQ:HLCS), Cambridge, Mass.
Life Technologies Corp. (NASDAQ:LIFE), Carlsbad, Calif.
Monsanto Co. (NYSE:MON), St. Louis, Mo.
Pacific Biosciences Inc., Menlo Park, Calif.