Tech News

Decoding protein structure, one femtosecond at a time

You know those high-speed cameras on the television show *Mythbusters*, the ones the hosts use to slow a bullet to a crawl or catch an explosive shock wave in mid-flight? Imagine using a camera like that to capture molecular dynamics. You could catch the motion of enzymes as they flex to bind ligands, watch photoreactive proteins change shape in response to light, or map a virus's topology.

Turns out, imaging at that level isn't science fiction: The first such "high-speed" instrument was installed near Stanford University in 2009 and has been pumping out a steady stream of images since 2011.

But it's not exactly a camera; the Linac Coherent Light Source (LCLS) is actually a laser, the world's first "hard" X-ray free electron laser (XFEL). There is a second XFEL, the SPring-8 Angstrom Compact free electron laser (SACLA) in Japan, and a third is under construction in Germany. Though specifications vary, these lasers all fire blindingly bright, short-wavelength X-ray light into samples in pulses just femtoseconds (10^{-15} seconds) long. Like the flashes of a stroboscopic camera, researchers have used these pulses to develop an entirely new paradigm in protein structural biology: serial femtosecond nanocrystallography, or SFX for short.

Let's talk about "SFX"

X-ray crystallography isn't exactly new—of the 96,980 structures currently stored in the Protein Data Bank (PDB), nearly

In 2011, a team led by John Spence was the first to implement the SFX technique.

The Coherent X-ray Imaging (CXI) experimental station at SLAC's Linear Coherent Light Source X-ray laser, site of most SFX successes to date. (SLAC Multimedia Communications)

86,000 were solved by X-ray crystallography. Clearly, this is an imaging method that works—so, why is there need for a new one?

The reality is, crystallography suffers from a number of drawbacks. First, it requires a crystal, and a big one at that, typically at least 0.1 mm long. But not every protein crystallizes well, especially large proteins or those residing in membranes, which limits the technique's reach. Second, X-ray radiation damages crystals as it images them, causing the images to blur. To circumvent that problem, the crystals typically are chilled in liquid nitrogen, but resolution still inevitably suffers. Finally, crystals are, by definition, static snapshots of a single protein configuration. Researchers interested in protein dynamics typically use NMR for their structural studies, and indeed, most of the non-X-ray structures in the PDB were acquired by that method.

SFX addresses all of these issues, says John Spence, Regents Professor at Arizona State University (ASU) and Director of Science at the NSF BioXFEL Science and Technology Center, who with ASU colleague Petra Fromme and Henry Chapman from the Center for Free-Electron Laser Science in Hamburg, led the team that first implemented the technique in 2011 to produce an 8.5 Å image of photo-

system I. According to Spence, the first advantage of SFX is that it can "outrun" radiation damage.

LCLS pulses come 120 times per second, but each pulse is just 50 femtoseconds long. To put that number into context, says Spence, the time it takes for an atom to vibrate once at room temperature is 100 femtoseconds, and the time it takes an electron to orbit once around an atom is about 1 femtosecond.

The energy delivered in those pulses is staggering. According to Janos Hajdu, Professor of Cell and Molecular Biology at Uppsala University, his team has achieved power densities in excess of 10^{20} W/cm² in a tightly focused XFEL beam. "If you focused all sunshine hitting the Earth to a millimeter-square, you get the same power density that one would get if you focused a pulse from the LCLS to 1 μ m² during the duration of the pulse," Hajdu says. The peak brilliance of the beam, he adds, is 9 to 10 orders of magnitude greater than synchrotron radiation, the current crystallography standard.

Because those pulses are so short and so bright—each contains some 10^{12} photons—the method is able to produce a sharp image before the crystal itself is damaged. A fraction of a second later, the crystal explodes, but not before the detector records the diffraction pattern. "We call

The site of the European XFEL facility, which lies as much as 38 meters below ground near Hamburg, Germany.

[that] ‘diffract-and-destroy’, or ‘diffract-then-destroy,’ Spence says.

That speed and brightness means SFX can handle crystals far smaller than a synchrotron can. Small crystals are easier to grow as a rule; they also are generally cleaner, as larger crystals typically contain more irregularities in the lattice structure. “You go down in crystal size by several orders of magnitude,” says Petra Fromme, Professor of Chemistry and Biochemistry at ASU. In one recent study to determine the 2.1 Å structure of *Trypanosoma brucei* cathepsin B, a potential drug target for African sleeping sickness, the crystals measured just $0.9 \times 0.9 \times 11 \mu\text{m}^3$. Some researchers, like Hajdu, are working to push the technique to the single-molecule level, and laser intensity on that scale is within reach. “We would need another factor of 10 to 30 or so,” he says.

Another advantage of SFX is that, because imaging occurs prior to radiation damage, crystals can be shot at room temperature, a more physiological condition. And then there’s the issue of dynamics. Conformational changes induced by ligand binding, for instance, can damage or destroy crystals, making imaging problematic—but not if you’ve got a never-ending supply of fresh material. “The shutter speed is fast enough to freeze the motion of atoms,” Spence says. One recent study, coauthored by Fromme, demonstrated the feasibility of capturing diffraction patterns from a cyanobacterial photosystem I complex 0, 5, and 10 microseconds after photoactivation, fast enough to catch glimpses of the profound molecular motions that occur over that period.

Not a crystal paradise

Still, SFX poses significant technical challenges. In traditional crystallography, single crystals are imaged repeatedly for extended periods. In SFX, the crystal is destroyed as soon as it’s hit. Thus, a steady stream of tiny crystals must be sprayed into the laser’s path, as if by a tiny atomizer, and

thousands of crystals are required for each angle—enough to capture every possible X-ray reflection many times over.

Spence, who with Bruce Doak and Uwe Weierstall developed the injectors, says they employ a thin stream of liquid that is compressed to micron size by a flowing stream of gas within a glass capillary. In this way, a 40 micron wide stream is squeezed into a beam just 1 micron wide, flowing at $10 \mu\text{l}$ per minute. Yet because the laser pulses, there’s considerable waste in the process. Vadim Cherezov, Associate Professor in the Department of Integrative Structural and Computational Biology at The Scripps Research Institute, says only one in “tens of thousands” of crystals is hit, meaning potentially hundreds of milligrams of protein are lost.

Experimental geometry for SFX at the LCLS. Single-pulse diffraction patterns from single crystals flowing in a liquid jet are recorded on a detector at the 120-Hz repetition rate of LCLS.

Recently, Weierstall designed a new injector system for membrane proteins, which Spence calls a “toothpaste jet.” This injector uses a gel-like substance called “lipidic cubic phase” (LCP), which “mimics the environment of the native membranes in which these receptors reside,” according to Cherezov, and also encourages their crystallization. Instead of jetting the material into the laser path, LCP is extruded slowly, like toothpaste from a tube. “It has the consistency of car grease,” Spence says.

For a recent paper using SFX to determine the structure of the serotonin 5-HT_{2B} G-protein-coupled receptor, Cherezov says he was able to reduce protein consumption “by about two to three orders

of magnitude.” In total, his team needed just 300 μg of purified protein. As an added bonus, those data were collected at room temperature. Comparison with a structure collected under cryocooled conditions in a synchrotron reveals some of the protein’s loops are in fact less rigid than they first appeared in their frozen state, an observation that has implications for protein dynamics studies, Cherezov says.

Life in the cave

To run SFX experiments, users can travel to the LCLS at Stanford, or to the SACLA in Japan. If all goes well, by 2017 they could instead head to the European XFEL facility in Hamburg, Germany.

These instruments are not identical, though all can be used for SFX. LCLS fires 50 femtosecond pulses at 120 Hz, while SACLA fires at 60 Hz. Hamburg’s XFEL is expected to have a pulse rate of 27,000 Hz, enough to complete experiments in a fraction of the time compared to LCLS and SACLA, and will feature 6 experimental stations, 3 of which can be used simultaneously.

According to Massimo Altarelli, Chairman of the Management Board at the European XFEL facility, the €1.3 billion project contains 5.8 km of tunnels extending from Hamburg northwest to the neighboring town of Schenefeld. It includes a 2 km superconducting linear accelerator, a fan of 5 undulator tunnels, and an experimental hall measuring 50 m $\times$ 90 m. The overall distance from the start of the linear accelerator to the experimental hall is 3.4 km.

The process starts in the linear accelerator, which produces a high-energy electron beam. The accelerated electrons [with an energy of 17.5 billion electron volts (GeV) at the European XFEL, slightly more than LCLS’s 13.6 GeV and twice SACLA’s 8 GeV] then hit one of the instrument’s undulator tunnels, “an absolutely fundamental part” of an XFEL, Altarelli explains. Undulators are arrays of magnets that force the electron beam into a zigzag, “like a skier who is doing a slalom course.” As the electrons turn, they emit X-rays, and if the undulator “is sufficiently long and sufficiently perfect,” Altarelli says, the result is a coherent beam, i.e., a laser.

Researchers cannot simply show up to use the instrument; time is awarded based on scientific merit. At LCLS, about one in five proposals is typically accepted, according to staff scientist Marc Messerschmidt, and each user is awarded between 3 and 5 12-hour blocks, generally once per day.

Spence describes the process of data collection as “living underground in a cave eating bad food for a week with jet-lag.”

“The students are all exhausted. You’re working 24 hours a day. It’s eight-hour shifts typically. So it’s like sailing a yacht across the North Sea ... and a third of the crew is asleep at any one time, another third is getting fed, and another third is on duty.”

First, users have to assemble the “kit” they need to do their experiments—the injectors, pumping systems, samples, and so on. Some researchers bring crystals, others prepare them in wet-labs on site.

Spence recalls that when he and Fromme were preparing the photosystem I crystals for their first SFX paper, the crystals had to be grown in the dark. “So Petra Fromme, she puts up a tent and you go inside this dark tent. And there’s this brew of green goo that’s growing these tiny little crystals, and they’re so small you can’t even see them.”

The injector system, especially, is still very much a work-in-progress, says Sébastien Boutet, a staff scientist at LCLS. “They’re not plug-and-play,” and they can clog. Researchers generally show up a few days early to test their equipment. When it’s actually time to run the experiment, somebody has to sit and watch the diffraction patterns coming off the detector to make sure the X-ray beam actually hits the sample. With both the beam and the liquid jet having micron-scale cross-sections, even the smallest vibrations can knock them out of alignment.

“Somebody stares at a screen for 12 hours and steers the jet back into the stream for 12 hours. If things go well, it’s very boring,” Boutet says.

Data, data, and more data

At LCLS, data are collected on CSPAD, the Cornell-SLAC pixel array detector, which consists of 64 detector tiles arrayed around a central hole through which the X-ray beam passes.

The detector captures 120 images per second, each 10 megabytes in size.

That’s about a gigabyte per second—comparable to the ATLAS detector at the European Council for Nuclear Research (known as CERN), according to Filipe Maia, Assistant Professor of Molecular Biophysics at Uppsala University. Typical data sets, Maia says, are on the order of 100 terabytes. Cherezov’s recent serotonin receptor study netted 4,217,508 images in 10 hours—about 42 terabytes total. That’s a bit much to fit on a USB key, or even a hard disk, so LCLS users must FTP the data home, and that itself can take a month.

Just storing the data is a burden. At Uppsala, Maia has about a petabyte (1000 TB) of storage, the vast majority of which is dedicated to XFEL data. He processes it on a cluster comprising 34 nodes of four graphical processing units each. GPUs excel at the Fourier transforms required to turn diffraction data into images. “We’re bound by the I/O and not so much by the computing,” he says.

Firing at 27,000 Hz, the European XFEL will present an even greater data challenge. For instance, according to Altarelli, the instrument will not fire at a constant rate. Rather, those pulses will come in 10 bursts of 2700 Hz. The European XFEL Facility has spent €20–30 million on detector design to tackle the problem, he says, and they’re still not quite there. To date, they can collect images at 3,000–5,000 Hz, “which is still quite a remarkable achievement,” he says.

To process those images, most SFX researchers use CrystFEL, a software suite developed by Thomas White, a scientist at the Center for Free-Electron Laser Science in Hamburg.

According to White, CrystFEL has to handle some unique challenges. First, instead of one crystal hit at defined angles, an SFX data set comprises thousands upon thousands of images, each with a random orientation that must be individually determined. The

detectors are still developmental, and beam intensity fluctuates wildly from pattern to pattern. “Obviously, this [processing] needs automation,” White says.

First, the data set is culled of empty images, which usually cuts its size down by an order of magnitude. Then, images are indexed—that is, the orientations of the crystals are determined—and the intensities of the spots (Bragg peaks) measured and combined. The output, White says, “is a merged data set of reciprocal lattice intensities,” that can be imported into standard structure determination software. For their serotonin receptor structure, Cherezov’s team collected 4.2 million images, of which 152,651 were actually crystal strikes. Of those, 32,819 were indexed (using the cleverly named *indexamajig* CrystFEL module) to produce the final 2.8 Å structure.

First, though, researchers needed to solve the so-called “phase problem.” As Cherezov explains, X-rays are a form of light, which has both amplitude and phase. Amplitude information is captured by the detector, but phase information is lost, and without that, it is impossible to transmogrify diffraction data into a structure.

To solve the phase problem, crystallographers use a number of tricks. Most SFX structures, including Cherezov’s GPCR, were solved by molecular replacement, which uses an already known structure as a kind of template to make the problem easier. Recently, though, researchers successfully determined the first de novo crystal structure by SFX, that of the model enzyme lysozyme in complex with gadolinium. The very dense lanthanide atom provides a reference point in the diffraction pattern to solve the phase directly.

“We’ve shown for the first time that you can actually determine the phases using free-electron laser data only,” says Thomas Barends, staff scientist in the Coherent Diffractive Imaging Group at the Max Planck Institute for Medical Research in Heidelberg, Germany, who was the lead investigator on that study. “That’s the crucial point: It means you can solve a protein structure without any prior knowledge about that structure.”

The question is, which proteins will the SFX community turn its focus to next?

Written by Jeffrey M. Perkel, Ph.D.

BioTechniques 56:55–58 (February 2014)

MagSi-DNA clean^{FIX}

Cut the cost of DNA clean-up: PCR & Sequencing from the same kit

- Robust high throughput protocols
- Replaces Agencourt CleanSEQ[®]
- Delivers high quality sequence data
- 99% removal of primers and other contaminants

amsbio