



## Structural biology shapes up

Despite advances in the field of proteomics, protein folding still remains a mystery. Yet innovations in X-ray crystallography, electron microscopy, and data analysis (think robots and Google) are yielding answers about protein structures faster than ever before. **By Alan Dove**

**W**ithin a single generation, researchers discovered that DNA was the genetic material encoding RNA, which in turn encodes proteins, and these proteins carry out the principal enzymatic reactions of life. They established that the sequences of proteins ultimately determine their 3D structures, which determine their functions. All that was left was to find a straightforward way to sequence genomes, predict the protein structures they encode, and build accurate models of all of life.

Half a century later, a few problems persist. One of the toughest is the challenge of getting from an amino acid sequence to a 3D protein structure. Though the former ostensibly determines the latter, the exact rules of the peptide-folding process remain largely inscrutable. Nonetheless, computational biologists continue to chip away at the problem, producing an evolving series of tools that can predict some classes of protein structures quite accurately. In the meantime, the classic technique of X-ray crystallography has become more accessible to nonspecialists, and a revolutionary series of developments in electron microscopy is revealing structures that previously eluded understanding.

### Advancing crystallography

Rather than trying to predict a molecule's shape from first principles, X-ray crystallographers attack the problem from the opposite direction, purifying and crystallizing a protein and then measuring how it diffracts a beam of X-rays. The technique can be tedious, often requiring thousands of experiments to determine the conditions that will yield usable

protein crystals. Today, though, biologists can let robots do much of the work.

"Robots enable one to set up much smaller crystallization drops, so that you need smaller amounts of sample to screen large numbers of conditions," says Cynthia Wolberger, a professor of biophysical chemistry at **Johns Hopkins School of Medicine** in Baltimore, Maryland. Manufacturers now offer robotic systems optimized specifically for testing crystallization conditions, such as the **TTP Labtech** mosquito and the **Formulatrix** Rock Imager. Meanwhile, rapid cloning and protein expression platforms let researchers produce multiple variants of a protein to find one that will crystallize well.

Crystallographers have also improved their ability to study membrane proteins, which have been notoriously hard to crystallize. By mixing lipids, water, and proteins in specific proportions, researchers can form a cubic liquid crystal amenable to X-ray diffraction. The method requires only nanogram quantities of protein. "This is a huge advance," says Martin Caffrey, a professor of biochemistry and immunology at **Trinity College** in Dublin, Ireland. Caffrey adds that studies of G-protein coupled receptors and other critical membrane proteins "really underwent an explosion as a result of this cubic phase methodology."

Once a protein crystallizes, investigators take it to one of a few government-funded synchrotron facilities to have it bombarded with X-rays and to collect their data. That process has also gotten easier in recent years. "We rarely go [to the synchrotron] anymore; we ship our crystals there," says Wolberger, adding that "it's all been set up with great software and robotics so you can operate it from anywhere."

Improvements in synchrotron data collection systems now enable analyses that would have been impossible just a few years ago. A process called "raster scanning" allows researchers to examine crystals—and small portions of crystals—to find areas with the best diffraction in a sample that would otherwise be unusable. Automated, high-speed data collection also permits scanning many more crystals. For example, Wolberger and her colleagues recently analyzed over a thousand crystals to gather enough information for a structure, an approach that would have been impractical with manual exposures.

### Upcoming Features

Genomics—October 7 ■ Neurotechniques—November 4 ■ Cell Analysis—November 25

### “Hit and run”

The X-ray beams themselves have improved too, with newer systems yielding higher levels of X-ray flux that enable faster, higher-resolution analyses. Meanwhile, a new technology called the “free-electron laser” has captured the attention of many structural biologists. “These are producing pulses of X-rays that are on the order of 10 to 40 femtoseconds long, so these are extremely short pulses, but there’s a huge number of X-rays in each one,” Caffrey explains.

That enables an entirely new type of crystallography, which Caffrey refers to as “hit and run.” Researchers hit a tiny crystal with the laser’s powerful X-ray pulse and collect the resulting data before the crystal explodes. While synchrotrons could already examine crystals as small as a micrometer, the new technique could enable analyses on crystals measured in nanometers.

As full-time crystallographers continue to push the limits of the technique, it’s also become more accessible to researchers in other fields. That’s especially true for labs working on relatively simple, soluble molecules, a description that encompasses the majority of the human genome’s protein products. “I teach courses where we explain how to crystallize proteins, [and if] you attend one of these you should be able to go home and set up crystallization trials,” says Caffrey.

Several vendors, including **Hampton Research**, **MiTeGen**, and **Molecular Dimensions** also sell ready-made protein crystallization kits, further enabling beginners to try their hand at the field. Those who succeed will find plenty of help with the next step. “There are people at synchrotrons who would be delighted to take on projects and to work with somebody at a noncrystallographic lab, and to solve a structure and become a coauthor,” says Caffrey.

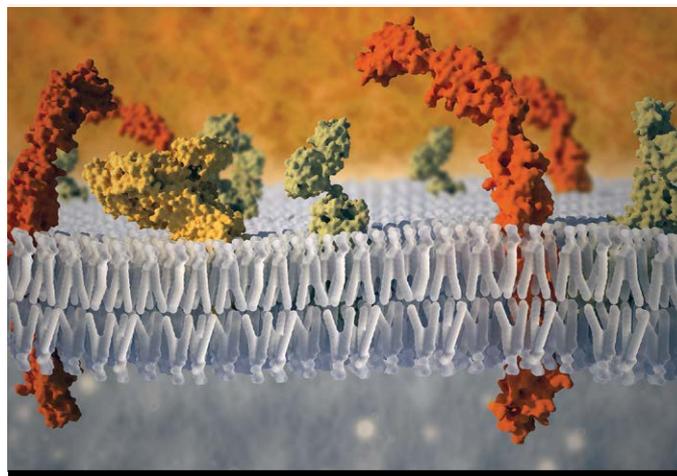
### Cold hard data

While X-ray crystallography has undergone a steady evolution, the other major protein structure determination technique, cryo-electron microscopy (cryo-EM), is in the midst of a revolution. In cryo-EM, researchers freeze the specimen of interest into a thin block of ice, then place it into an electron microscope, photograph it, and analyze the image to determine the target’s structure. Though it’s long been a useful strategy for studying relatively large structures such as viruses, technical problems limited its utility.

That suddenly changed a few years ago, thanks partly to advances that had nothing to do with the microscope itself. “By far the most dramatic [development] is new detector technology,” says Eva Nogales, a professor of biochemistry and molecular biology at the **University of California, Berkeley**. Electron microscopists traditionally relied on film to capture their data, but that method came with a host of limitations. Film canisters only held 50 frames, setting a maximum number of exposures in a single sampling session, and processing, scanning, and analyzing the analog images could introduce errors.

Engineers eventually designed sensitive, radiation-hardened digital sensors that could replace film even in high-energy electron microscopes. The new sensors detect electrons directly, providing extremely high resolution with none of the inconveniences of film. “The contrast was really excellent, so we had much higher signal in those images, and they came with an extra bonus: They have very, very fast readout,” says Nogales.

That development helped address another longstanding problem. The electron beam causes the frozen sample block



“It’s a fabulous time in the field, and there’s tremendous excitement.”

— David Agard

to warp, motion-blurring a traditional camera’s images. Cryo-EM researchers previously dealt with that by scanning through numerous images to find the few clear ones taken when the protein wasn’t moving. With the new sensors, “the readouts are fast enough that you can take a movie during that exposure and then realign the sample to correct for that beam-induced motion, and that’s had a huge impact,” says David Agard, a professor of biochemistry and biophysics at the **University of California, San Francisco**.

### Two angstroms and beyond

The latest generation of sensors, developed with funding from the U.S. government’s economic stimulus program after the 2008 recession, can detect individual electrons. That enables cryo-EM to yield high-resolution protein structures with molecular weights of just a few hundred kilodaltons, previously the exclusive domain of X-ray crystallographers. Electron microscopists are cautiously optimistic about their ability to push further into crystallography’s turf. “Whether we are going to be able to get to resolutions that are sometimes achievable by X-ray crystallography, like two [angstroms] and beyond, I do not know,” says Nogales, adding that “right now the record [for cryo-EM] is around 2.2 angstroms for one very well-behaved sample.”

Indeed, even dedicated cryo-EM labs still see a need for crystallography. “If you can get crystals, for the most part you get a structure immediately with crystallography, so there’s no reason not to try to do that; in my lab we do that as a matter of course,” says Agard. However, he adds that “there’s going to be a whole realm of things that haven’t crystallized well that will be taken over by cryo-EM.”

Software that runs cryo-EM systems and processes their data has also improved dramatically, making everything from microscope alignment to data mining faster and easier. While the software is mostly open source, the hardware certainly isn’t. Modern cryo-EM microscopes come from a single **cont.>**

## FEATURED PARTICIPANTS

**Centre for Integrative Systems  
Biology and Bioinformatics**  
www.imperial.ac.uk/integrative-  
systems-biology

**Computation Institute,  
University of Chicago**  
www.ci.uchicago.edu

**Direct Electron**  
www.directelectron.com

**FEI**  
www.fei.com

**Formulatrix**  
www.formulatrix.com

**Gatan**  
www.gatan.com

**Hampton Research**  
www.hamptonresearch.com

**Johns Hopkins School  
of Medicine**  
www.hopkinsmedicine.org

**MiTeGen**  
www.mitegen.com

**Molecular Dimensions**  
www.moleculardimensions.com

**Trinity College Dublin**  
www.tcd.ie



**TTP Labtech**  
www.ttplabtech.com

**University of California,  
Berkeley**  
www.berkeley.edu

**University of California, San  
Francisco**  
www.ucsf.edu

**University of Chicago**  
www.uchicago.edu

ADDITIONAL  
RESOURCES

**CAMEO**  
www.cameo3d.org

**Phyre2**  
www.sbg.bio.ic.ac.uk/phyre2

**RaptorX**  
raptorx.uchicago.edu

company, **FEI**, and sensors for them are made by it, **Direct Electron**, and **Gatan**. Regardless of which vendors and instruments one chooses, a complete cryo-EM system costs several million dollars to buy, and several hundred thousand annually to maintain and operate. “Only the richest labs or institutions that are subsidizing this will be able to do this, and that’s not a good situation in terms of national capabilities,” says Agard. Germany, the United Kingdom, and China have all established national-level funding for cryo-EM facilities, while the U.S. National Institutes of Health is still deciding how to proceed.

**Siri, what does the protein look like?**

Regardless of how fast X-ray crystallography and cryo-EM develop, though, neither is likely to satisfy researchers’ growing desires for structural models. “You can’t beat an experimentally determined structure, [but] the challenge is the vast number of genomes that have been sequenced, and the effectively impossible task of devising 3D structures for all the proteins,” says Michael Sternberg, director of the **Centre for Integrative Systems Biology and Bioinformatics** at Imperial College London, United Kingdom.

To address that challenge, Sternberg and his colleague Lawrence Kelley have focused on the decades-old dream of molecular biologists: predicting protein structures directly from sequence information. The team’s latest tool, **Phyre2**, uses template-based modeling, a process that compares a sequence against a global database of previously solved protein structures. Several other research groups have developed similar tools, which are accessible online to researchers worldwide.

These tools can yield impressive results, at least for protein families that are well-represented in the databases. “On

some protein samples we can do pretty well, since now we have much more data, so we can apply or develop much more sophisticated techniques to model the sequence–structure relationship,” says Jinbo Xu, a senior fellow of the **Computation Institute** at the University of Chicago. Xu is the principal investigator behind **RaptorX**, another online portal for predicting secondary and tertiary protein structures.

In addition to template-based modeling, some researchers are now exploring a technique called “contact prediction.” This approach searches through vast troves of sequence data to identify evolutionarily conserved amino acid interactions, then uses those correlations to predict a novel sequence’s folding patterns. “That’s certainly proving very useful for membrane-bound proteins where there are very few crystal structures available,” says Sternberg, but he adds that “you still need quite a large number of aligned sequences” for contact prediction to work.

Besides improving the underlying algorithms, computational biologists have been working on making their system interfaces friendlier. “We’re very much following the sort of Google approach of a very clean screen, not inundating the user with many options, and really only delivering what the user wants,” says Sternberg. It seems to be working. Sternberg estimates that of the 44,000 unique users who accessed **Phyre2** last year, only a few hundred contacted him or Kelley for support.

**Expanding options**

Xu and his colleagues have also embraced the user-friendly model, with similar results. “The broader community is using the tools, [and] the users of my server have very diverse backgrounds,” he says. Researchers without structural biology training may not understand the limitations of the underlying algorithms, so most of the major portals are designed to help users interpret their results. **RaptorX** provides a quality evaluation along with each protein model, scoring how likely the structure is to be correct. Similarly, **Phyre2** provides an overall confidence score, as well as individual scores for substructures down to the amino acid level.

Because all of the major structure prediction tools are free online, scientists can also hedge their bets by sending their target sequences to all of them to see how the models differ. Another site, **CAMEO**, also tests and rates the different protein-structure services. Each week, **CAMEO** sends a test sequence to all of the participating servers, then calculates benchmarks based on the systems’ speed, server reliability, and other characteristics.

Different tools also have distinct strengths and weaknesses depending on a researcher’s specific needs. **RaptorX** emphasizes template-based modeling for particularly challenging proteins, while other portals are better suited for rapid analysis of large numbers of relatively well-characterized protein families. **Phyre2** offers a suite of additional services; the “**Phyre Alarm**,” for example, will send an email to researchers when new data allow the system to calculate an improved model of their protein.

Regardless of the approach they choose, investigators seeking protein structures should expect more good news. Surveying the progress in all three techniques, Agard echoes the general sentiment of structural biologists: “It’s a fabulous time in the field, and there’s tremendous excitement.”

*Alan Dove is a science writer and editor based in Massachusetts.*

DOI: 10.1126/science.opms.p1600107