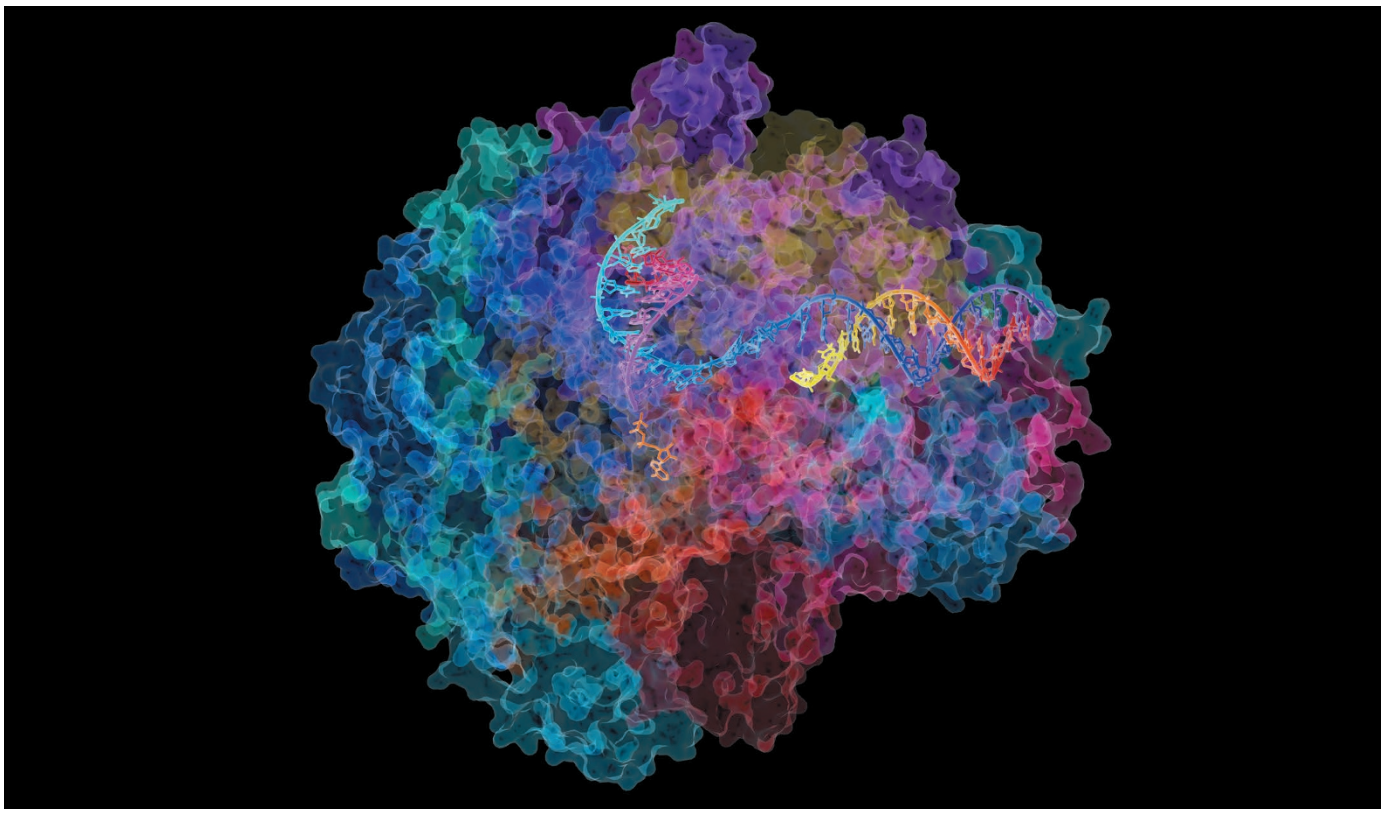
TECHNOLOGY FEATURE

LET THE STRUCTURAL SYMPHONY BEGIN

Structural biologists are at last living the dream of visualizing macromolecules to uncover their function. But it means integrating different technologies, and that's no easy feat.

DAVID BUSHNELL, KEN WESTOVER & ROGER KORNBERG/STANFORD UNIV.



The details of the enzyme RNA polymerase have been intriguing structural biologists for some time.

BY STEPHEN ORNES

Like other structural biologists, Eva Nogales works in extraordinary times. The University of California, Berkeley, faculty member now has the tools to tackle important questions about cells' molecular machinery that would have been impossible to answer just a few years ago.

A recent project with Berkeley colleague Jennifer Doudna, the molecular biologist who co-pioneered the CRISPR–Cas9 gene-editing method, is a case in point. Both were intensely interested in the R-loop, a structure made of nucleic acids that forms in cells in many situations, but also just before DNA is snipped by CRISPR–Cas9. Nogales and her team revealed an R-loop in *Streptococcus pyogenes* bacteria, and from the near-atomic-resolution images, deduced how the Cas9 enzyme opens up the DNA conformation at specific sites and makes them accessible to CRISPR's molecular scissors[1].

The work is remarkable for the speed with which the scientists assigned a function to the structure, but also because they arrived at the solution by combining imaging methods — an increasingly popular approach in structural biology. For more than a century, the field's premiere method has been X-ray crystallography. But some biomolecules are simply too big or small to crystallize, and the technique doesn't work on others. And some biomolecules change shape or orientation as they work, which isn't captured by static crystallization.

Now, scientists have a dazzling suite of different imaging techniques with which to build on crystallographic findings. Some of the approaches, such as cryogenic electron microscopy (cryo-EM) or chemists' stalwart nuclear magnetic resonance (NMR) imaging, reveal molecular shapes, size and orientation at near-atom-level resolution without the need to make crystals. But not every method works for every protein, nucleic acid or other biomolecule inside a living cell.

Growing wisdom in the field suggests that no single method is likely to be sufficient to probe the dynamic behaviour or intricate interactions taking place in a cell. The most powerful insights will come from hybrid ▶

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.

methodologies that integrate the images from several different tools.

The approach is rapidly gaining followers. "Each [method] brings something important to the table, and the combination is very much larger than the sum of the parts," structural biologist Roger Kornberg of Stanford University in California. Kornberg won the 2006 Nobel Prize in Chemistry for his work detailing the machinery of gene transcription. For that ground-breaking research, he generated crystallographic pictures. Now, like other crystallographers, he has moved on to hybrid methodologies.

Kornberg continues to analyse RNA polymerase II, but now he combines crystallography with cryo-EM, in which an electron beam probes the structure of biomolecules. Cryo-EM can be used on molecules that don't crystallize easily and can reveal larger structures than can X-ray crystallography, but — for the moment at least — it lacks crystallography's high resolution. Kornberg's lab also uses chemical crosslinking and mass spectroscopy to reveal relationships between nearby proteins, and homology modelling to construct representations using information from known proteins[2].

Nogales and Doudna's team also took the hybrid route to study R-loops. "The full R-loop could not be seen by the high-resolution X-ray crystallographic structure," says Nogales. So they also used cryo-EM to reveal the full R-loop structure at lower resolution. Only by combining the two methods could the researchers work out how R-loops fit into the larger CRISPR–Cas9 picture[1].

Such hybrid, or integrative, approaches help researchers to probe deep basic-science questions, but also reveal details that are useful to drug developers. Large proteins found in cell membranes are often targets for therapeutic drugs, and high-resolution hybrid methods have the potential to show in atomic detail how a drug interacts with a receptor. Similarly, hybrid methods might be able to aid vaccine development by showing how proteins on the viral envelope of HIV, Ebola and other pathogens interact with immune cells to induce protective responses. "These structures are super-important to understand how our immune system works," says structural biologist Jens Meiler at Vanderbilt University in Nashville, Tennessee.

Nogales sums it up: "This is a golden time to do hybrid methodologies."

LIVING THE DREAM

The current era in structural biology promises to fulfil the "dream of many life scientists", says Jan Ellenberg, head of the Cell Biology and Biophysics Unit at the European Molecular Biology Laboratory (EMBL) in Heidelberg, Germany. That dream is to seamlessly scale up from what scientists see at the atomic level to the cellular level. Such deep understanding of the cell's macromolecules naturally leads to answers to the overarching question in structural biology — how is a molecule's structure connected to its function?

Each technique in a structural biologist's toolbox offers a different perspective. Models that use hybrid methods can boost biologists' confidence that a model accurately reflects how the molecule or ensemble acts in the cell. "You need all of them in combination to really get a full understanding of your biological question of interest," says Meiler.

"One of the really exciting things about cryo-EM is that you can start a biochemical process and freeze those samples at multiple states."

X-ray crystallography has long reigned as the standard way to determine the atomic structure of proteins. Of the 120,000 or so models in the Protein Data Bank (PDB), established in 1971, about 90% were derived from crystallographic studies.

But structural biologists' workhorse, even with its high resolution, has limitations. Crystallography requires highly purified samples that produce a well-ordered crystal. Scientists fire X-rays at a crystal to determine its structure by analysing how the atoms scatter light. The technique needs a specimen with enough atoms to produce a measurable diffraction pattern, and every crystal must be static. As a result, the method can't reveal how a molecule moves or functions in a cell, or its connections to other systems.

A protein "is not just a single static structure", says Gunnar Schröder, who leads the computational structural biology group at the Institute of Complex Systems in Jülich, Germany. "Oftentimes, what you want is to see how the whole protein works." Schröder uses hybrid methods to understand the movements and connections of proteins. Crystallography provides a snapshot of a protein in one configuration, removed from its normal environment. He says that structural biologists need other methods to boost the structural information from crystallography and improve their understanding of the form and function of proteins.

Many proteins, such as drug targets on a cell membrane, are flexible and often unstable. To get these proteins to form crystals, researchers often have to change them in some way. Meiler says the altered specimen may not accurately reflect the native state of the molecule or how it is arranged in the cell. He mixes experimental and computational approaches to better understand molecular structure. "It takes time for people to understand that for many biological systems, the model from crystallography is a good starting point," he says, but it may not be suitable for providing information about function.

Biologists are now leveraging a range of tools to build richer, more accurate models of biological structures. Hybrid approaches have the power to do more than a single technique ever could. One particularly useful partnership joins cryo-EM and X-ray crystallography. This microscopy method has been around since the 1980s, but in recent years it has achieved a resolution of 2.2 ångströms, edging close to the 2 Å average resolution of X-ray crystallography[3]. It can produce models in two or three dimensions of proteins and other macromolecules that have stubbornly resisted other approaches.

"One of the really exciting things about cryo-EM is that you can start a biochemical process and freeze those samples at multiple states," says Jeffrey Lengyel, principal scientist for life sciences at FEI, a company in Hillsboro, Oregon, that designs and manufactures cryo-electron microscopes. "You can determine the structure of multiple conformations."

Researchers can also combine cryo-EM images to see molecules in motion. John Rubinstein of the Hospital for Sick Children in Toronto, Canada, led work published in May 2015 that used image analysis to combine 100,000 cryo-EM images into a film, showing changes in the structure of eukarotic V-ATPase, an enzyme that pumps proteins across membranes and, changes over time[4]. In papers published earlier this year, Nogales and her collaborators used cryo-EM with homology models to describe the structure of TFIID, a large, horseshoe-shaped protein complex that is required to initiate gene transcription[5].

WITH A LITTLE HELP

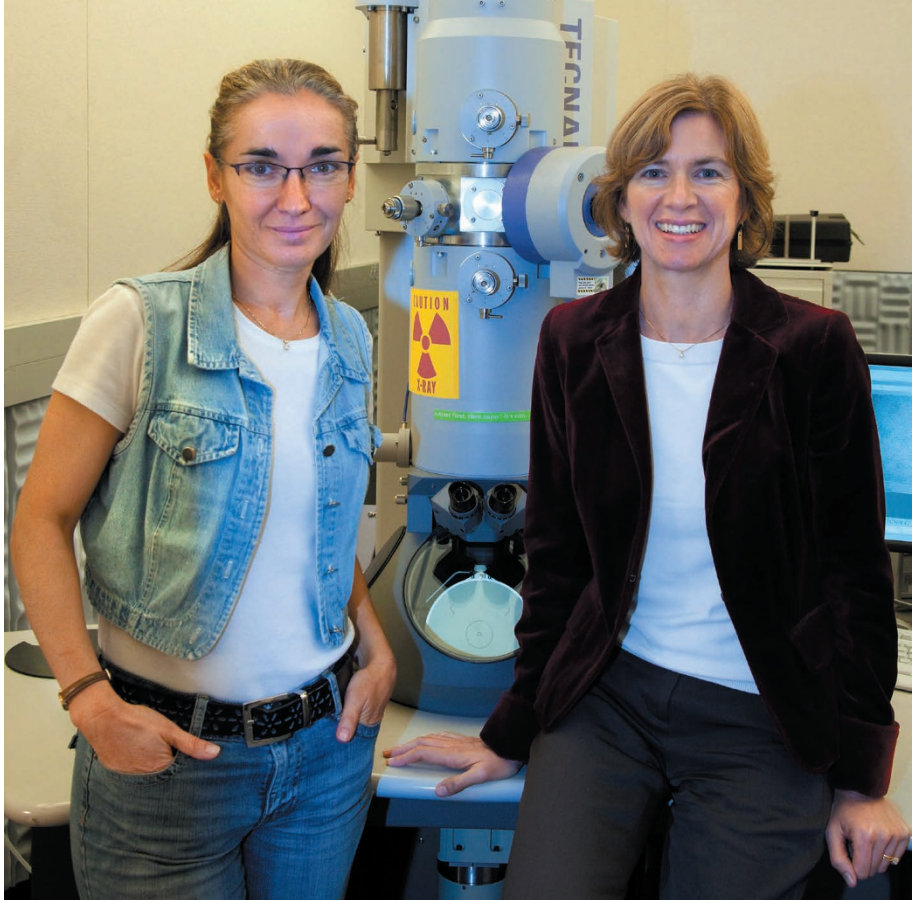
The hybrid strategy of Nogales and her team led to an overall resolution of better than 10 Å — a significant improvement over their previous analysis of the same protein at 30 Å. That resolution has led to new insights: "We can see what amino acids are interacting with DNA," Nogales says.

But cryo-EM requires specimens to be snap-frozen. That's not ideal for biological samples, as the conditions are far removed from a macromolecule's dynamic, natural state. NMR spectroscopy can help on that front. "NMR has a big advantage in that you can look at proteins at room temperature, and get information on dynamics," says Schröder, whose lab builds experimental models that combine NMR data with those from cryo-EM and crystallography.

First used experimentally in the 1940s, NMR reveals macromolecular structures by exciting atoms in an external magnetic field. When the atoms relax, the changes in their internal magnetic fields can be mapped to each atom. However, NMR spectroscopy works only on relatively small macromolecules or ensembles.

Structural biologists are also using hybrid

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.



LAWRENCE BERKELEY NATL LAB.

Eva Nogales (left) and Jennifer Doudna worked together to reveal how the Cas9 enzyme uncoils DNA in preparation for gene editing.

methods to tackle supersized ensembles, a task that would have been impossible in the past. Kornberg's latest research, which has not yet been published, extends his ongoing RNA polymerase II studies and uses hybrid methodologies to describe a giant assembly made of more than 50 proteins and transcription factors. "The entire assembly could now be visualized for the first time through the combination, and I would say equal contribution, of all the methods," he says.

Another supersized target is the nuclear pore complex. This collection of membrane proteins acts as a gatekeeper for information and molecules passing in and out of the nucleus. In 2015, Ellenberg and his colleagues used a hybrid approach to study the structure of this protein behemoth[6]. In the past, researchers had probed the complex with crystallography and electron microscopy, but they weren't able to image the entire thing at molecular resolution, and its overall structure largely remained a mystery.

Ellenberg's team first imaged the nuclear-pore complex using super-resolution fluorescence microscopy, which he says can identify features measuring less than 30 nanometres. To improve the resolution, they combined it with an image-processing technique called single-particle averaging that uses information from thousands of pores, bringing down the resolution to about 10 Å. Comparisons with cryo-EM maps of the same complex validated their work. The result is a zoomed-in view of a supersized protein complex. The EMBL team "generated models of the nuclear pore that were unthinkable in the past", Nogales says.

Similarly, Rubinstein and Lewis Kay at the University of Toronto used hybrid methods to push the boundaries of what was deemed possible. By combining cryo-EM with NMR spectroscopy, they mapped previously unidentified conformational changes of an enzyme called VAT, which has an important role in breaking down proteins in a cell. Cryo-EM revealed the structure, and used together with NMR, they were able to show how the enzyme changes shape, painting an elegant portrait of a protein at work[7].

HYBRID DRAWBACKS

Although biologists are gaining clarity from merging different tools, each technique also contributes its own error rates. Mixing them, therefore, presents a potential problem because it multiplies the sources of error. "How can I combine these different ways of analysing error into one holistic approach that gives me a measure of confidence, accuracy and precision in model?" asks Meiler.

Yet another hurdle is melding different data sets to make them accessible and useful to other researchers. The rich level of information from any one technique makes this a formidable challenge. "You can literally generate terabytes of data per day," says Lengyel. He hopes the structural-biology community might benefit from the approaches in astronomy and high-throughput genetics to grapple with data overload. Although software exists that can neatly combine high-resolution crystallographic data into cryo-EM maps, other hybrid methodologies aren't as straightforward to merge. Electron paramagnetic resonance spectroscopy, for example, measures distances and orientation in a macromolecule, whereas cryo-EM produces a density map. Although those two measurements would be useful together, they don't speak the same language. "How do I combine these very different metrics? How do I share these data?" asks Meiler.

To discuss the best ways to organize, share and use data from hybrid approaches, dozens of structural biologists gathered in October 2014 at the European Bioinformatics Institute in Hinxton, UK[8]. The meeting was the first of its kind, organized by a task force set up by the Worldwide Protein Data Bank. At present, the PDB stores data from individual protein structures, says Schröder. "We should get to the point where we have all the information that we know about this protein — all the different conformations it can take," he says. Such rich data, he says, will help to reveal the bigger picture of proteins and other big molecules.

There have been steps in that direction: archives exist for electron-microscopy models in two dimensions (the Electron Microscopy Pilot Image Archive) and three dimensions (EMDataBank). Established with funding from the EMBL and other sources, these archives contain data that can be shared, archived and distributed.

Yet another challenge threatens to forestall progress in the field: human expertise. "Investments are needed in technology, but it's equally important to invest in educating scientists," says Meiler. He recommends that students learn the limitations and challenges of each method — and become an expert in at least one. "We need to train a new generation of scientists who are capable of understanding how to integrate these different technologies," he says.

Finally, structural biologists have to learn to ask new, complicated biological questions that may seem impossible. Thanks to hybrid methods, says Ellenberg, "things have come within reach that even five years ago I wouldn't have been able to dream of doing until retirement". ■

Stephen Ornes *is a science writer in Nashville, Tennessee.*

1. Jiang, F. *et al. Science* **351,** 867–871 (2016).
2. Murakami, K. *et al. Proc. Natl Acad. Sci. USA* **112,** 13543–13548 (2015).
3. Bartesaghi, A. *et al. Science* **348,** 1147–1151 (2015).
4. Zhao, J., Benlekbir, S. & Rubinstein, J. L. *Nature* **521,** 241–245 (2015).
5. Louder, R. K. *et al. Nature* **531,** 604–609 (2016).
6. Szymborska, A. *et al. Science* **341,** 655–658 (2013).
7. Huang, R. *et al. Proc. Natl Acad. Sci. USA* **113,** E4190–E4199 (2016).
8. Sali, A. *et al. Structure* **23,** 1156–1167 (2015).

© 2016 Macmillan Publishers Limited, part of Springer Nature. All rights reserved.