STRUCTURES OF DESIRE

What do protein crystallographers dream of? The eukaryotic ribosome, the spliceosome, the nuclear-pore complex, the HIV trimer and almost any transmembrane protein, finds **Ananyo Bhattacharya**.

When considered up close, the blood protein from a sperm whale is a marvellous thing. Or so it seemed just over 50 years ago, when John Kendrew and other researchers at the Cavendish Laboratory in Cambridge, UK, reported that they had used X-rays to reveal the three-dimensional structure of a globular protein for the first time. Analysis of the diffraction pattern caused by crystals of myoglobin, which was chosen for its simplicity, required one of the most powerful computers in the world at that time, and later won Kendrew a share of the Nobel Prize in Chemistry with his Cavendish colleague Max Perutz. The picture it created “is more complicated than has been predicated by any theory of protein structure”, Kendrew and his colleagues wrote in a *Nature* article¹.

Half a century on, X-ray crystallography's techniques are in outline the same: you need a crystal, X-rays and calculating power to make sense of the diffraction pattern. In all these three areas, however, progress has been

enormous. Washing machines now have more computing power than the computers Kendrew used and synchrotrons offer X-rays a trillion times more brilliant than those available in Cambridge 50 years ago. Growing crystals is increasingly routine: with many proteins it can be an automatic process, and the skills needed for the harder ones have come on in leaps and bounds.

But so, too, have the ambitions of the crystallographers. They still believe, as did those earliest molecular biologists, that there is no better way to understand a complicated machine than to capture an atomic-scale picture of it. But now they dream of using their techniques on things far more challenging and complex than anyone imagined 50 years ago — including some machineries of protein and nucleic acid that dwarf myoglobin as a whale does a minnow.

THE BUILDING SITE

The crystals look beautiful under the microscope. Suspended in a tiny drop of solvent, each gem-like growth is a symmetrical array of copies of one of the two subunits of **the ribosome** — the cell's machine for making proteins. Each of those subunits is a tangle of many proteins and RNA. What makes the crystals especially interesting is that their constituents come not from bacteria but from eukaryotic cells. Which plant, animal or other eukaryote provided those cells, though, is for the moment a secret.

Such crystals are rare and highly sought after. Unfortunately, Nenad Ban ruefully admits, these ones grown in his lab at the Swiss Federal Institute of Technology in Zurich (see photo, right) will not be yielding a protein structure any time soon. A crystal that looks pretty in visible light can still be a mess by the demanding standards of X-rays. “We have fantastic-looking crystals,” he says. “But they are not very good in terms of diffraction.”

Ban has been here before. Nearly a decade ago he was in one of several teams striving to

solve the structure of the simpler ribosomes used by bacteria and archaea. He and his colleagues in Tom Steitz's lab at Yale University published the structure of the 50S subunit of the ribosome of *Haloarcula marismortui* — a microorganism that lives in the famously salty waters of the Dead Sea — in 2000 (see ref. 2). The 50S subunit consists of about 30 proteins and has a mass of 1.5 million daltons, compared with myoglobin's 1,700 daltons or a carbon atom's lowly 12. Just a month later, a second team published the structure of the smaller 30S ribosomal subunit, using material from the bacterium *Thermus thermophilus*³. The following year, Harry Noller at the University of California, Santa Cruz, and his colleagues unveiled the structure of the whole bacterial ribosome, revealing much about how it binds to the transfer RNAs that deliver amino acids to a growing protein⁴.

These structures triggered an avalanche of new work in a field that had, in their absence, largely ground to a halt. They allowed researchers to see, for example, how the ribosome

catalyses the joining of one amino acid to another with a peptide bond. Now, says Jennifer Doudna of the University of California, Berkeley, “we know a lot of detail about what the bacterial ribosome looks like, how it works, how peptide bonds are made, and even a lot about how the initiation process is regulated in bacteria”.

The eukaryotic ribosome takes the competition to another level. It is bigger — containing some 80 component proteins compared with the bacterium's 50 to 60 — but that is not the only, or even the main, challenge. “There's a lot more bells and whistles; there's a lot more regulation that goes on,” Doudna says.

It's those bells and whistles that make the eukaryotic ribosome more difficult to work with than its bacterial counterpart. In mammals, for example, a host of additional proteins called initiation factors interact with the ribosome. The initiation factors are themselves

THE EDITING SUITE

Noller, who with colleagues cracked the structure of the complete bacterial ribosome, says the next big thing is another signature speciality of the eukaryote, **the spliceosome**. “The spliceosome would be fantastic,” he says. “That would make the ribosome look like child’s play.”

Made up of 150 or so proteins, the spliceosome slices and dices freshly made messenger RNA, stitching together the ‘exon’ sequences that will be translated into protein and relegating the ‘introns’ to the cell’s cutting-room floor. It can bring together sections of RNA maybe tens of thousands of base pairs apart and then snip out the intervening loop, like a movie editor running through many metres of film to splice two shots together. The activity of the spliceosome fascinates biologists because it could help explain how eukaryotic cells generate biological complexity — in the form of different RNAs and proteins — from a single DNA sequence. But that fast and continuous activity is also what makes it a gruelling challenge for crystallographers.

“Tremendous work has been done genetically and biochemically to understand how splicing works and how it’s regulated,” says Doudna. “The big missing piece in that field is not having access to high-resolution structural information for how the spliceosome is put together and what is driving the ... changes that have to occur during the splicing process.”

The problem for crystallographers is that the spliceosome is not just one machine — it is five, and all are in constant motion. Called small nuclear ribonucleoprotein particles (snRNPs, pronounced ‘snurps’), these five protein–RNA complexes come together transiently in a complicated, fast-moving dance, their

mercurial assembly about the size of a ribosome but far less stable. “Unlike the ribosome, you cannot simply purify a spliceosome from cellular extracts because you have many different spliceosomal complexes — different snapshots at different stages of function,” explains Reinhard Lührmann of the Max Planck Institute for Biophysical Chemistry in Göttingen, Germany, whose group has crystallized several proteins that are components of snRNPs.

Stalling the spliceosome at a particular stage of the cycle is a “major challenge” says Lührmann. One way might be to use a mutant RNA message or a small molecule to arrest the spliceosome mid-splice.

A team led by Kiyoshi Nagai at the Medical Research Council Laboratory of Molecular Biology in Cambridge, UK, recently published a relatively low-resolution version of the crystal structure of one of the smaller snRNPs, which they reconstituted from its RNA and seven recombinant proteins (see graphic below)⁵. Reconstituting all 150 proteins in the complete spliceosome, however, “does not appear feasible” in the foreseeable future, Lührmann says.

Instead, he is pursuing the structure using cryo electron microscopy, in which samples are flash-frozen in liquid ethane to protect them from the bombardment of high-energy electrons in an electron microscope. Recent advances in this technique have allowed it

to show structures at resolutions as low as 5 angstroms, but do not yet approach the 2-angstrom resolution of good X-ray crystallography. Lührmann predicts that further advances in hardware and software will allow cryo electron microscopy to fill out the broad atomic structure of the spliceosome “in the next few years”.

M. MUELLER/THE BAN LAB

REF. 5

complex assemblies — eukaryotic initiation factor 3 (eIF3), for example, is made up of at least 12 proteins and is only a few times smaller than the ribosome itself. Ribosomes that are purified from a cell could be in any number of combinations with these and other proteins.

Getting material that is pure and homogeneous is a significant hurdle, says Doudna. Simpler proteins can be mass-produced by inserting the appropriate gene into a cell culture that then churns out proteins, but ribosomes are too large and complex to be produced in this way.

Some who worked on the bacterial structures are joining the hunt for the eukaryotic one, and the competition is fierce. That is why Ban will not reveal from which eukaryote his ribosomes are harvested, nor which subunit his lab has managed to crystallize. The fact that he has crystals — albeit ones that don’t diffract X-rays well — is an important proof of principle, he says. But “it’s the endgame that counts”.

THE MONSTROUS MAW

More than 30 times the mass of the ribosome and around 100 nanometres wide, **the nuclear-pore complex** is a doughnut-shaped assembly of hundreds of proteins that straddles the eukaryotic nuclear membrane. One of the largest protein conglomerations in the cell, the structure serves as both gate and gatekeeper, choosing which nucleic acids, proteins and other molecules to let in and out of the nucleus.

With only around 200 pores in a yeast cell —

compared with perhaps 10,000 to 20,000 ribosomes in a bacterium — purifying the complex from cells is “really impossible”, says André Hoelz, a researcher who works on the nuclear-pore complex in Günter Blobel’s laboratory at The Rockefeller University in New York.

Hoelz and his colleagues are taking a different tack — expressing and crystallizing single proteins or small protein complexes from the pore, and then piecing them together like a

jigsaw puzzle to reconstitute the whole structure. "When we started this work five years ago people were saying that you can't possibly get this done because of the sheer size of the nuclear-pore complex," Hoelz says.

There are two factors working in the group's favour. First, the nuclear-pore complex has an eightfold rotational symmetry and twofold mirror symmetry (see graphic). That means that the hundreds of proteins that make up the pore are in fact made up of repetitive arrangements of only about 30 different types — around half the variety in the ribosome. Second, in many organisms the nuclear-pore complex is dismantled when cells break down their nuclear membrane before dividing, and later reassembled piece by piece. It is always broken into the same building blocks, and it is these conserved components that Hoelz hopes to crystallize and slot together. If this approach is successful, it will provide a detailed 'pseudo-atomic' picture of the nuclear pore's structure, although it may not provide the clarity that a crystal structure of the entire complex would.

There's another feature of the nuclear-pore complex that is a challenge for crystallographers. The centre of the pore is a mesh of fluttering protein filaments that don't fold in a regular way, but instead flop and dangle; they are thought to play a crucial role in selecting which proteins are ferried through the pore. Because a crystal structure is really an average of the arrangements of the atoms in millions of protein molecules in the crystal, the tentacles, which are in constant motion, would be an ill-defined blur. "Imagine one could crystallize the nuclear-pore complex; a quarter of it would be natively unfolded, and that's the business end," says Michael Rout, whose Rockefeller University lab also studies nuclear-pore complexes. So far, only isolated structures for some bits of the tentacles exist, solved by teams that break them off and crystallize them separately.

The nuclear-pore complex, then, runs up against a fundamental limit of crystallography — it generates snapshots, not movies. And it is not alone: up to 30% of eukaryote proteins are wholly or partly disordered. To see proteins in action, some crystallographers and modellers have turned to computer simulations that jump between two or more 'frames', each obtained by crystallography. Wayne Hendrickson of Columbia University in New York says there is also a lot of excitement about technologies that might be possible at facilities such as the European XFEL (X-Ray Free Electron Laser) under construction

in Hamburg, Germany (see page 16). The idea here is that an extremely short burst of X-rays could be scattered off a single protein molecule, blowing it apart but revealing something about its structure before the disintegration. "You are in principle able to capture the molecule in action," Hendrickson says.

That's something for the future: the XFEL will not be commissioned until 2014. For now, Rout says, "the current approach is probably the most successful, which is to continue to crystallize the structured parts and put that together with other data to build a complete picture".

THE KILLER KEY

"A big missing piece in the virology field is the structure of **the HIV trimer**," says Ian Wilson, a structural biologist at the Scripps Research Institute in La Jolla, California. "We don't understand what that looks like." It's a challenge that Wilson and his colleague Robert Pejchal have taken on relatively recently — though they're not the first. "There have been many people in and out of the game for years because it has been so challenging."

The trimers that Wilson and others want are protrusions from the surface of HIV, also called 'envelope spikes'. Each one has a tripartite structure: three gp41 proteins rise from the viral envelope, forming a stem that supports three gp120 molecules. It has long been thought that when the spike binds to key receptors on white blood cells it triggers massive structural changes in the trimer that drive the fusion of virus and cell.

Trimers are the focus of intense vaccine research efforts, but vaccines based on the trimer so far do not stimulate enough of an antibody response to combat a later infection (see *Nature* 454, 565–569; 2008). "There's something about the trimer that makes it difficult to mount an effective immune response," Wilson says. If they had the trimer's structure to work from, researchers hope they could devise better ways to turn the human immune system against it.

The problem has been that, isolated from the virus, the trimer falls apart. Researchers can't adopt the nuclear-pore tactic here and break it apart before piecing it back together. The situation is more like that of the ribosome, in which the whole structure is more than the sum of its parts. Some regions of the trimer — those that are important for the virus's interactions with cell-surface receptors and for generating an effective immune response — are buried inside or in the interface between its parts.

Researchers can't tell what they normally look like unless they see a trimer intact.

The first structure of a pruned version of gp120, combined with a fragment of the white cells' CD4 receptor and an antibody, was published in 1998⁶. Other structures of gp120 and gp41 have followed^{7–9}. And a recent cryo-electron-microscopy study of whole virus particles revealed some of the large-scale changes that occur when the trimer binds to the CD4 receptor¹⁰. But the structure of the trimer proper has remained out of reach.

Wilson is making progress in collaboration with John Moore at Weill Cornell Medical College, New York, who has engineered a disulphide bond between gp41 and gp120 that helps to lock them together. They also used a strain of the virus that tends to form more stable trimers — called SOSIP trimers — in the first place. "We have SOSIP trimers and we have crystals, but they don't diffract well," says Wilson.

Still, Wilson and others will continue to plug away at the problem — because a vaccine remains "one of the big challenges to science in the present day", says Hendrickson.

The gp120 protein (green), a cell-surface protein (blue) and an antibody (orange, yellow and red).

BEST OF THE REST: MORE STRUCTURES HOTLY DESIRED

Internal ribosome entry site

A sequence of RNA that allows ribosomes to begin protein production from the middle. Helps RNA-based viruses to hijack protein production.

The sequence is very flexible. Parts have been crystallized — but not the whole thing.

Sec61 (eukaryotic translocon)

A protein complex that can help

an unfolded polypeptide chain to cross or span a membrane. Recalcitrant when it comes to crystallization.

Telomerase RNP

Adds repeating sequences of DNA to the end of chromosomes and plays a role in cancer and ageing. Individual components of the telomerase have been solved.

It's a flexible, hard-to-get

material and it interacts with other molecules.

Higher-order chromatin structure

The complex of protein and RNA with which DNA is packaged. The core of a single nucleosome — the basic subunit of chromatin — has been solved to high resolution.

Crystallographers want higher-order structures of chromatin — two or preferably more

nucleosomes. This is proving difficult.

Sodium-selective ion channel

Ion channels that allow sodium ions to pass but not potassium ions — essential in generating electrical potentials in neurons.

They are formed by one big polypeptide chain, which may make them harder to crystallize than potassium channels.

THE INVISIBLE THREAD

Stephen White's website at the University of California, Irvine, features an exponential curve of which structural biologists are proud. In 1985, it shows, the first structure of a membrane protein was solved¹¹. (That work, on the photosynthetic reaction centre, won a Nobel prize.) Now, the number of structures for proteins that span membranes deposited in the Protein Data Bank has climbed to more than 180.

Some membrane proteins are refusing to join the cavalcade, however. Take, for example, **the epidermal growth factor receptor** — the target of Genentech's breast cancer drug Herceptin (trastuzumab). Despite a decade or more of intense study, only the bits protruding outside and inside the cell membrane have been crystallized. The connecting portion — the bit that spans the membrane and transmits information from one side to the other — has not. The same is true for the 60 or so other proteins in this family of receptor tyrosine kinases, which have central roles in cell proliferation, differentiation and disease. Solving this delicate stretch of protein would begin to explain how a signal outside the cell — such as a growth factor, hormone or other 'ligand' that binds to the receptor — can cause a change in protein conformation that leads to a response inside the cell.

To crystallize a membrane protein, you have to ease it out of its normal milieu. Released from their membranes, though, the proteins easily lose their shape and precipitate out of solution. So detergents are used to keep them soluble, folded and active, and these can end up being a problem themselves. Unless just the right types and amounts of detergents can be found, their molecules can obstruct the interactions between proteins that allow them to line up and form crystals.

The particular problem for the receptor tyrosine kinases is not their bulk or complexity but

their flimsiness. The peptide chain that connects their extracellular and intracellular parts snakes through the membrane only once. The head or tail can wobble around on the single transmembrane stem and make it difficult for the protein to form ordered crystals.

Persistence has paid off when it comes to another major class of membrane proteins.

There was much jubilation among structural biologists when, in 2000, a group led by Masashi Miyano at the Riken Harima Institute in Japan crystallized rhodopsin, a light-activated protein purified from the cow retina (see graphic, showing the helices of the protein embedded in a membrane)¹². It was the first structure to be resolved in the class of G-protein-coupled receptors, a family of membrane receptors with almost 1,000 members found in humans. Solving a second such receptor, which was done in 2007, was still a marathon task: one group set up 15,000 trials using a robot to optimize crystallization conditions^{13–15}. These proteins have seven membrane-spanning regions, and may have been more tractable because this wider 'bridge' between head and tail stops their extremities wobbling around so much.

There's a long way to go yet, however, according to Hendrickson. "From my perspective," he says, "these structures haven't answered the questions that I want to answer, which are about how the activation process happens — how these proteins do their job when activated by a ligand."

A full-length receptor tyrosine kinase remains a dream structure, and one that many crystallographers doubt can ever be realized because there is no obvious way to stabilize the head and tail. "I'm not certain that's going to be feasible," says Hendrickson.

White says that membrane proteins in general, though, are getting less intimidating. "There are a lot more people with the courage to tackle membrane protein," he says. Their courage will grow as ways to make better crystals and brighter X-rays come online.

At the cutting edge, however, where crystallographers face the complexity of a nuclear pore or the wavering heart of a transmembrane protein, something extra is needed. "Know the protein intimately," White says. "So far, that seems to be a really important issue — to have somebody who loves the protein," he says. ■

Ananyo Bhattacharya is *Nature's* deputy news editor.

1. Kendrew, J. C. *et al. Nature* **181**, 662–666 (1958).
2. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. *Science* **289**, 905–920 (2000).
3. Wimberly, B. T. *et al. Nature* **407**, 327–339 (2000).
4. Yusupov, M. M. *et al. Science* **292**, 883–896 (2001).
5. Pomeranz Krummel, D. A., Oubridge, C., Leung, A. K. W., Li, J. & Nagai, K. *Nature* **458**, 475–481 (2009).
6. Kwong, P. D. *et al. Nature* **393**, 648–659 (1998).
7. Zhou, T. *et al. Nature* **445**, 732–737 (2007).
8. Huang, C.-C. *et al. Science* **310**, 1025–1028 (2005).
9. Chan, D. C., Fass, D., Berger, J. M. & Kim, P. S. *Cell* **89**, 263–273 (1997).
10. Liu, J., Bartschaghi, A., Borgnia, M. J., Sapiro, G. & Subramaniam, S. *Nature* **455**, 109–113 (2008).
11. Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. *Nature* **318**, 618–624 (1985).
12. Palczewski, K. *et al. Science* **289**, 739–745 (2000).
13. Cherezov, V. *et al. Science* **318**, 1258–1265 (2007).
14. Rosenbaum, D. M. *et al. Science* **318**, 1266–1273 (2007).
15. Rasmussen, S. G. F. *et al. Nature* **450**, 383–387 (2007).