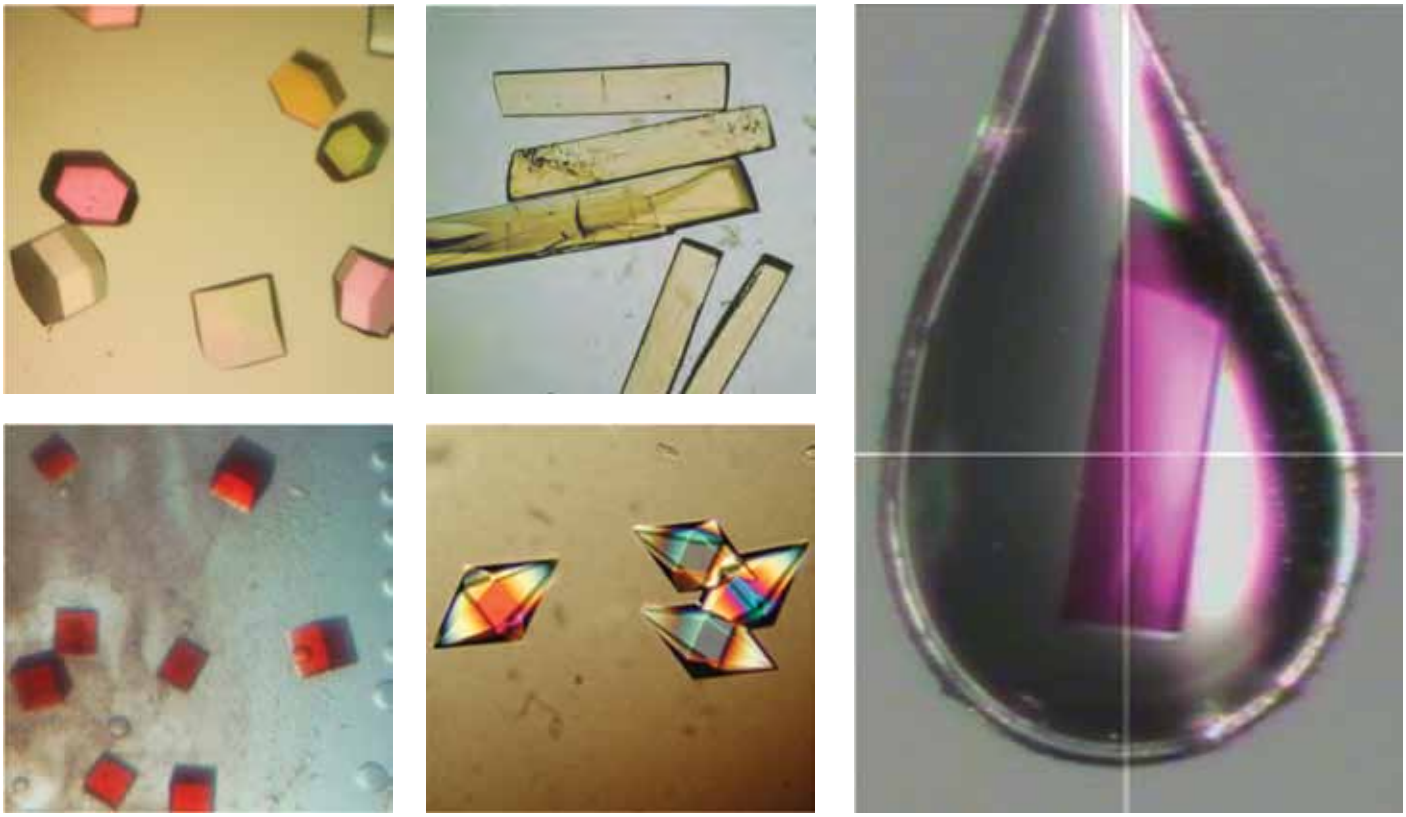


Protein crystals come in many shapes, sizes and colours. Most are colourless, but a cofactor or ligand (smaller chemical entities bound by proteins) may cause them to be coloured. Protein crystals are normally very soft and break easily. They need to stay in their 'mother liquor' to be able to diffract X-rays. Nowadays, most crystals are cooled to very low temperatures in nylon loops (right) to reduce damage from the X-rays. Images: Top left: http://web.chem.ucsb.edu/~kalju/chem112L/public/Crystals_11/crystals_S2.html Top right: http://iss.jaxa.jp/kibo/kibomefc/spcf_e.html Bottom left: http://www.bbsrc.ac.uk/news/industrial-biotechnology/2011/110524-pr-electricity-from-microbes.aspx Bottom right: http://www.usask.ca/research/100yrsinnovation/1_search_skies.php Large right: https://biocars.uchicago.edu/page/biology-customized-macromolecular-crystallography

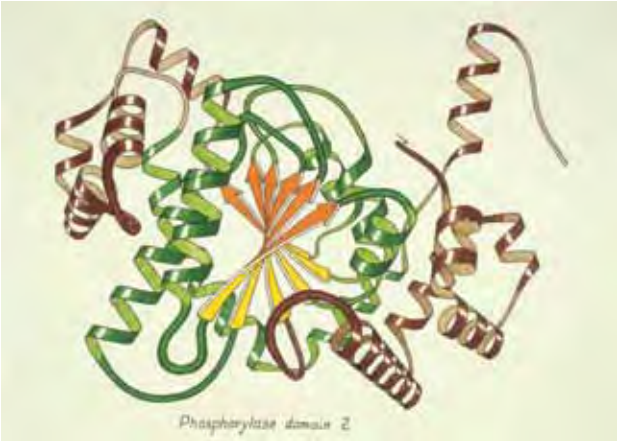
Seeing is believing

Wolf-Dieter Schubert explains the science behind macromolecular crystallography.

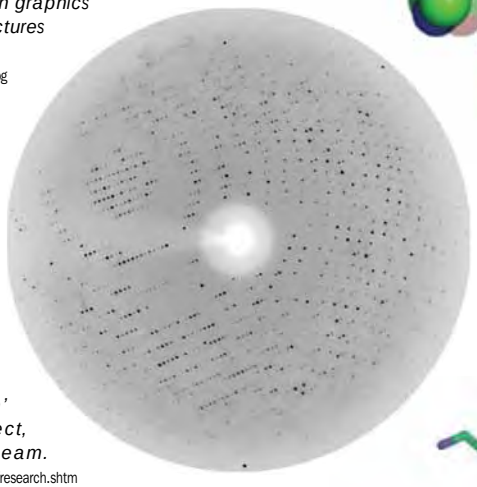
Have you ever wondered how something small like a virus or a bacterium can infect a person and cause anything from minor irritations such as a sore throat, to more severe symptoms like fever and measles, or even kill? Or have you marvelled at the elegant yet intricate double helix structure of DNA? How do we know that these things really exist or how they work? And how do we know what medicines to take to treat a particular disease? Thinking about or answering such questions was way beyond our reach until a few decades ago. Then came the discovery and refinement of crystallography in the first half of the 20th century and its application to biology from the 1950s onward. Slowly this opened our eyes to the wonders of the microcosm, showing us how biology works at the cellular, molecular and even atomic level.

We live in a physical, three-dimensional world. We are born with two eyes and two ears set slightly apart. This allows us to perceive the world around us. We not only see the height and the width of any object but its depth too. This in turn allows us to understand how one thing relates to another – is it smaller, thicker, broader, behind or in front of another? Are boxes neatly stacked one on top of the other or are they randomly thrown onto a big heap?

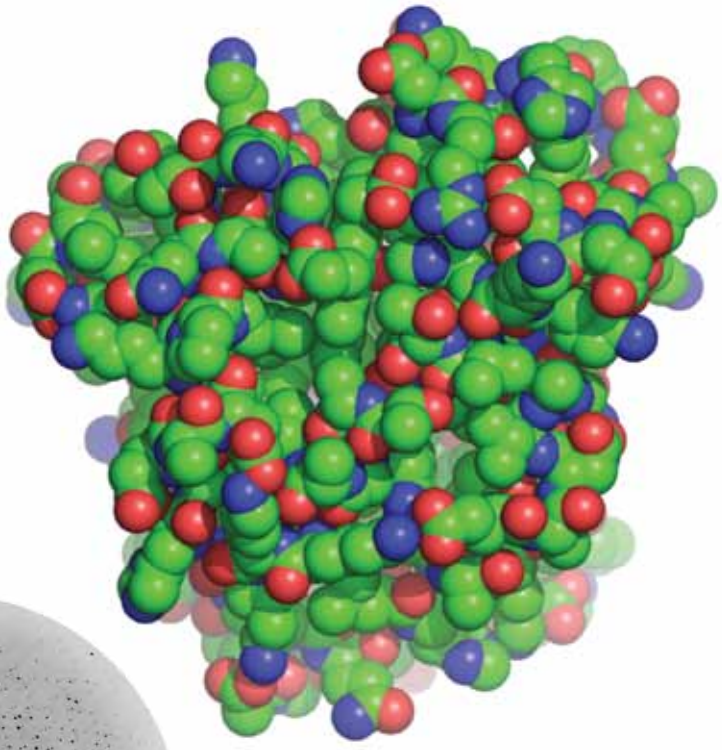
To analyse a protein normally requires milligram amounts of that protein. This may not sound a lot, but it is not easy to get. Generally bacteria are tricked into making the protein. They are grown in flasks with medium (yellow) at 37°C. Image: Structural Biology Initiative Cape Town 2010 (No longer exists



In the early days of macromolecular crystallography, representations of molecules had to be drawn by hand, as in this historic work of art of phosphorylase domain 2. Since then graphics programs have been improved allowing pictures to be prepared much more quickly.
Image: http://commons.wikimedia.org/wiki/File:Phosphorylase_domain2.jpg



Protein X-ray Diffraction: *Proteins are large molecules compared with salts and small chemicals. As a result X-ray diffraction images contain many reflections (black spots). To collect a complete data set, the crystal is rotated around an axis that will leave the crystal in the X-ray beam. Many such images are recorded and evaluated together. The white spot in the middle is caused by a 'beam stop' that protects the detector from the direct, high-intensity X-ray beam.*
Image: http://folk.uio.no/utek/research.shtm



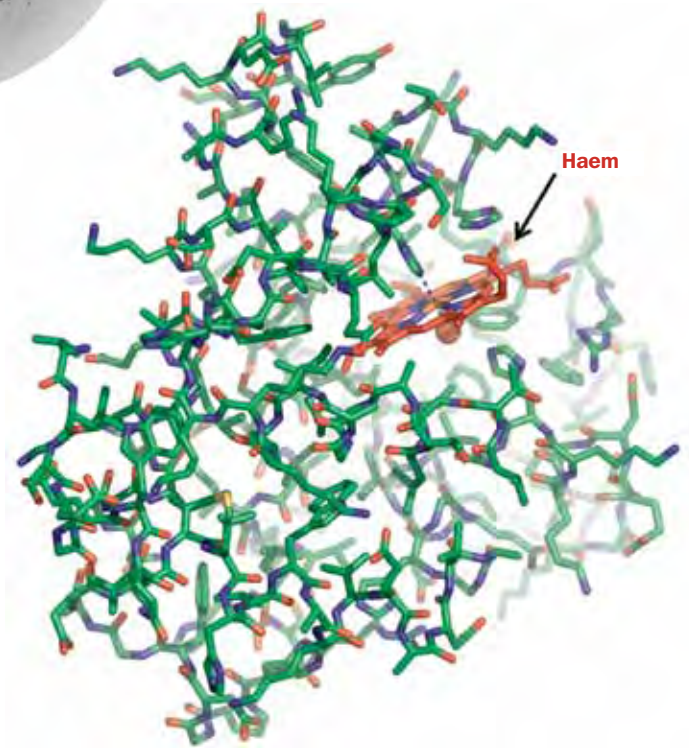
Myoglobin was the first macromolecule whose structure was experimentally determined. This was done in Cambridge, UK, in 1956 by John Kendrew and colleagues. Even for this fairly small protein it is difficult to see what is going on because there are too many atoms. Carbon – green, oxygen – red and nitrogen – blue. Image: PDB entry 1MBN in Pymol

In the same way, scientists would like to understand how microscopically small objects like molecules and viruses interact with each other. Fortunately, we can now do this for biological molecules with the help of the technique known as macromolecular crystallography.

Biological crystallography

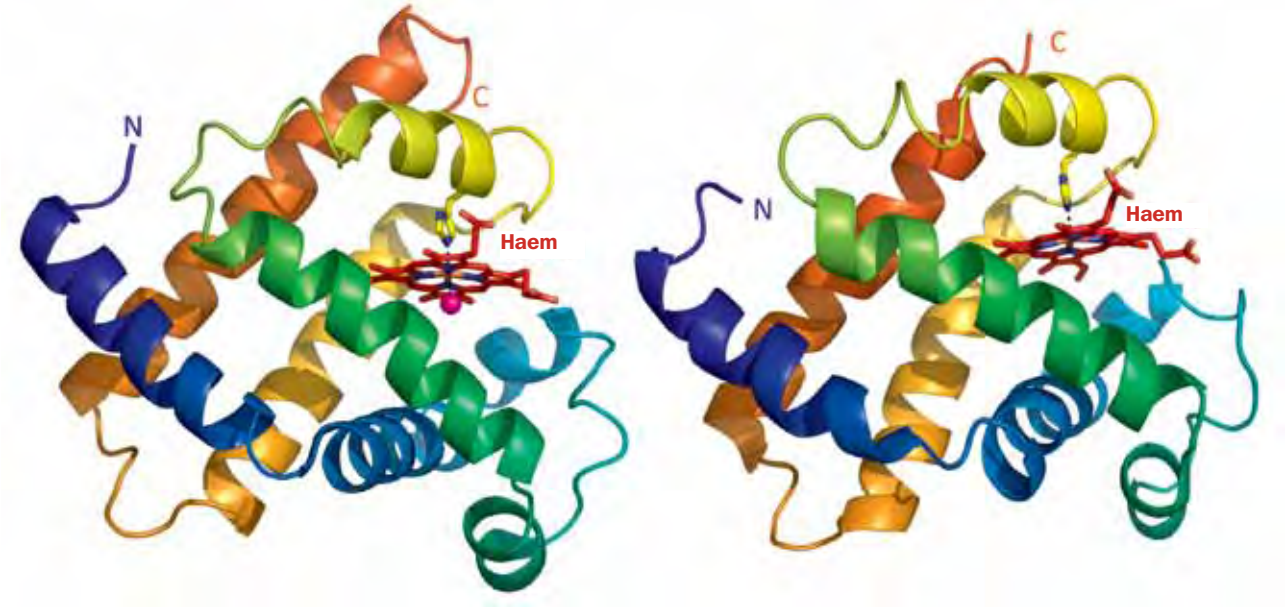
Exactly when biological crystallography started is open to debate. Many proteins crystallise spontaneously in their natural environment. One example is insulin, which regulates our blood sugar levels and which occurs as tiny crystals before being secreted into the blood stream. Another is alcohol oxidase, which crystallises in peroxisomes inside yeast cells grown on methanol. This allows dense packing of the protein without losing its activity. More frighteningly, protein crystals sometimes form due to genetic mutations. For example, crystalin, a protein found in the human eye, may crystallise and damage the eyes if certain mutations occur. Or haemoglobin may crystallise within red blood cells, preventing them from squeezing through narrow arteries and causing a lot of pain and suffering.

Scientists have been able to purify and crystallise proteins since around 1850. Protein crystallisation was seen as proof of protein purity by around 1900. Crystals are neat stacks of millions upon millions of molecules. The fact that proteins crystallise means that they have a defined three-dimensional shape. Before that proteins were assumed to be shapeless colloids similar to jelly. However, the technology to analyse



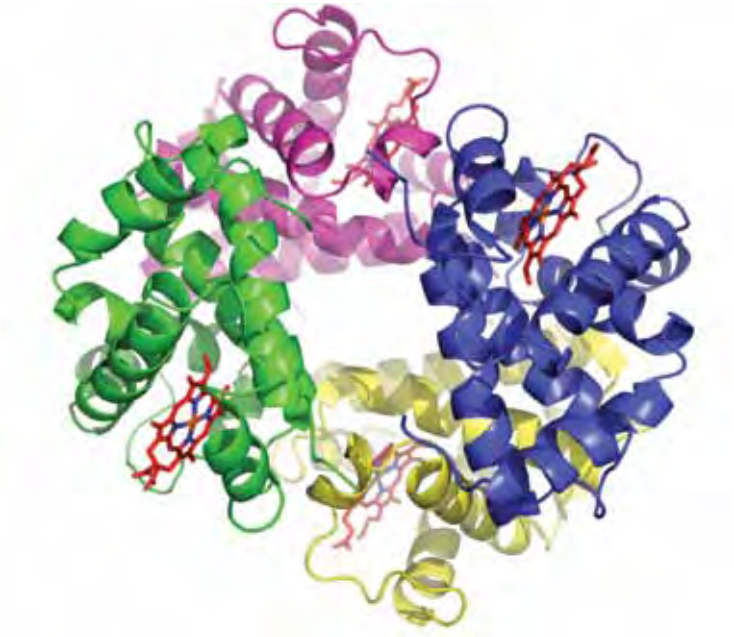
If we show only the chemical bonds connecting the atoms instead of atoms themselves, things are a little clearer. For example, we can now see where the haeme group is, but it is still too complicated. (Hydrogens are not shown for clarity.) Image: PDB entry 1MBN in Pymol

protein crystals was not yet available. This changed after the discovery of X-rays at the end of the 19th century, the realisation that X-rays could be used to probe the structure of crystals and many, many innovations producing better X-rays, better detectors and, of course, computer soft- and hardware.



We can take just one central atom from each amino acid making up a protein and connect these up. Now we can see that myoblogin starts at one point ('N' for amino-terminal end) and ends at another (marked 'C'). In between the chain forms spirals called a-helices, which in turn are connected 'loops'. Overall the chain packs up quite tightly, forming a pocket for the haeme cofactor (red), which binds the oxygen (pink sphere). Haeme is dark red in colour, giving blood its red colour.

***Haemoglobin:** Can you spot differences between myoglobin on the left and a single chain of haemoglobin on the right? Although the amino acid sequences are quite different, the fold is essentially the same.* Image: PDB entry 2DHB in Pymol



Haemoglobin actually consist of four chains. Each chain binds a haeme cofactor (red), allowing haemoglobin to carry four times as much oxygen. Image: PDB entry 2DHB in Pymol

The nature of enzymes

Until around 1930, it wasn't clear whether enzymes (catalysts of biochemical reactions) are proteins. Many people thought that enzymes were chemical things that could bind to proteins but could also leave again. By crystallising the first enzyme urease, which was still active inside the crystal in 1926, James Sumner proved that enzymes are indeed proteins. In fact most of the first protein structures to be solved were those of enzymes. In 1934, John Bernal and his 24-year-old student Dorothy Crowfoot (later Hodgkin) were the first people to demonstrate that protein crystals do scatter X-rays using crystals of pepsin, a stomach enzyme. The secret that had escaped everyone else was to keep the crystals moist and bathed in their 'mother liquor', the solution in which they crystallise, during the experiment.

Haemoglobin and myoglobin

The first protein structures derived from a full crystal analysis were those of haemoglobin and myoglobin, two molecules involved in oxygen transport and storage. These could be isolated in large quantities from blood and muscle respectively. First understood was the structure of smaller myoglobin solved by John Kendrew and colleagues in Cambridge, England.

Two years later the same group worked out a much more detailed myoglobin structure. Around the same time, Max Perutz and colleagues worked out a low-resolution haemoglobin structure. These first protein crystal structures were revolutionary, as we could for the very first time see molecules of life magnified to a level where we can see, feel and understand their properties. Unexpectedly, the fold of the protein chain in myoglobin turned out to be very similar to that of each of four chains in haemoglobin. Clearly the molecules are related, meaning that evolution also applies to molecules. In 1962 Max Perutz and John Kendrew were rewarded for their hard and pioneering work with the Nobel Prize for Chemistry.

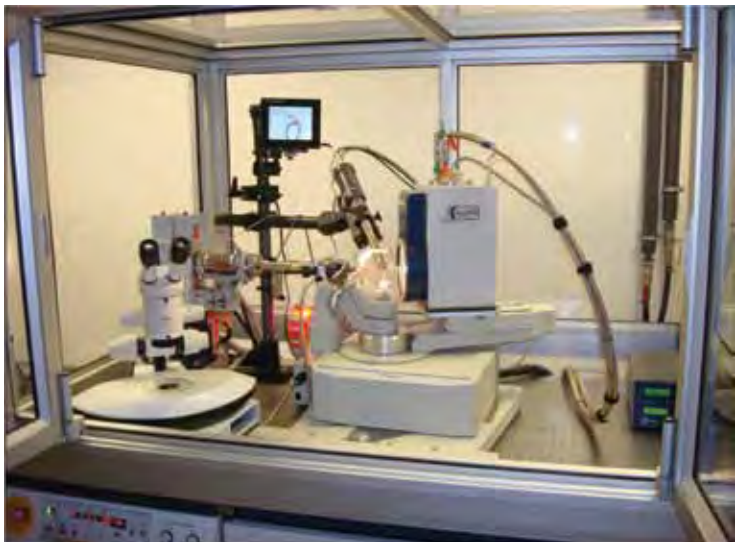
The structures of haemoglobin and myoglobin immediately made a lot of physical sense. Each protein is like a chain with twenty differently coloured beads or amino acids along the way. Each chain has a single starting point and a single end. In between, the chain folds in a complicated but unique way that leaves no gaps. In haemoglobin and myoglobin the chain wraps around another smaller entity called haeme, which is the actual oxygen carrier. The haeme cofactor is dark red in colour and the reason for that blood is red. Understanding the structure of haemoglobin and myoglobin also allowed the scientists to show that the protein chain is locally arranged as spirals called α-helices that had previously been predicted by Linus Pauling. A second way of efficiently packing proteins, discovered later, is by zigzagging the chain up and down, forming a sheet-like arrangement. These structures are called β-sheets. Individual β-strands are often depicted as arrows (top left figure on previous page).

A robot can be employed to optimise the use of purified protein during crystallisation. A robot can dispense much smaller amounts of protein than a human can do by hand.
Image: Structural Biology Initiative Cape Town 2010

More and more structures

The theories and the methods of solving protein crystal structures developed by Max Perutz, John Kendrew and colleagues allowed many scientists around the world to solve crystal structures of their favourite proteins – a revolution that is continuing to this day. At first enzymes were very popular, as methods had been developed by other scientists to isolate large amounts from the various organisms. As time went by, though, technologies for copying and changing DNA became more established. Now scientists could trick certain bacteria, yeast, plants or other cells into making large quantities of proteins, originally made in entirely different organisms.

Over the years, the size and the complexity of crystal structures have increased. First, mainly enzymes were analysed. These could reveal exactly how one molecule is changed into another with the help of the amino acids making up the protein chain. Then more complicated, multimeric proteins started being analysed. These proteins are composed of more than one chain and the arrangement can either be spherical, hollow spherical, tubular, cup-shaped and many other shapes. Virus coat structures are among the most complicated structures, made up of tens to hundreds of identical units. Quite a lot of these have been solved, telling us which part of the virus protein is likely to be responsible for infecting human cells.



X-ray diffractometer: Protein crystals diffract X-rays only weakly. As a result high-brilliance X-ray sources coupled to a very sensitive detector are needed to record the diffraction data. There are currently two such systems in South Africa – one in Cape Town and the other in Johannesburg.
Image: Structural Biology Initiative Cape Town 2010

European Synchrotron Radiation Facility (ESRF): Synchrotrons are very large, custom-built scientific instruments that produce high-intensity X-rays. Hundreds of scientists can simultaneously run many experiments in stations spread all around the ring. South Africa is a participant in the ESRF located in Grenoble, France, as there is currently no synchrotron in Africa.

The structure of ribosomes

Another major achievement was the determination of ribosome structure. Ribosomes are the protein-making factories in cells. The information from a particular gene is fed into the ribosome in the form of messenger RNA. The ribosome reads this genetic sequence and converts the information into a linear sequence of amino acids that then folds into an active protein doing its particular job. The structure of the ribosome allows us not only to think of the processes in broad brushstrokes, but to really understand it in a lot of detail. As a result we can understand why certain drugs work by binding to important places in the ribosome and either slowing or speeding up the process of making a protein. And ideally we can improve those drugs by looking how they bind and whether that binding can be improved.

Membrane structure

An important class of proteins are those that are embedded in membranes. Membranes are the outer perimeter of most cells. They are made up of particular fat molecules, scientifically called lipids. As fats and oils don't mix with

X-ray Free Electron Laser (XFEL): SACLA (SPring-8 Angstrom Compact Free Electron Laser) is one of only three X-ray free electron lasers in the world. It is adjacent to the SPring-8 synchrotron (circular building at the back) in Hyogo Prefecture, Japan (near Kobe/Osaka). Insert: Samples are destroyed by X-ray lasers within femtoseconds (millionth of a billionth of a second), but useful information is still obtained before this happens. Image: http://xfel.riken.jp

water, membranes form a barrier around the cell. Without proteins spanning the membrane, cells would die as very few substances would be able to get either in or out of a cell. Proteins form channels across the membrane, letting through only very specific ions or molecules which the cell needs. On the other hand, membrane proteins also act as the eyes and ears of cells. These proteins are called receptors. Receptors wait around for the right molecule to swim past the outside of the cell. They then grab hold of that molecule and either pass it directly to the other side of the membrane or send the message across the membrane that food, hormones or possibly a poisonous substance is around. The cell as a whole can then respond by either moving in the direction of the signal in the hope of finding more such molecules – or moving away from harm.

Improved technology

To be able to solve structures of complicated proteins and complexes, much experimental progress was needed over the years. This included better, faster and more automatic laboratory equipment, better crystallisation techniques and, very importantly, much brighter X-rays and much more sensitive detectors. From the 1970s onward, synchrotrons provided more and more intense X-ray beams, allowing smaller and smaller crystals to be used. Synchrotrons are essentially circular pipes with magnets around them. By synchronising the magnets, subatomic particles can be made to move around the synchrotron within the pipes at ever increasing speeds. These speeding particles produce high-intensity electromagnetic radiation including X-rays away from the circular pipe. The radiation can be made to pass through crystals which will then scatter the X-rays. By collecting information on the scattered X-rays, the structure of the molecule making up the crystal can be inferred.

More recently X-ray lasers have been built. At present there are only a few of these lasers in the world and people are still working out exactly what to do with them. X-ray lasers produce beams of extreme brightness, which destroy or evaporate anything in their way. But before evaporating, crystals still briefly scatter the X-rays, allowing crystal structure to be determined. The advantage is that these crystals can be really tiny, allowing much more difficult projects to be attempted. These difficult projects involve very delicate proteins, which generally don't crystallise easily at all.

Macromolecular crystallography has been around for just over 50 years. Many achievements have resulted in Nobel Prizes and countless other awards. Open any modern textbook on biochemistry, microbiology, molecular and cellular biology and you will see numerous pictures describing what molecules look like and how their shape and arrangement of amino acids allows them to do whatever they are meant to. The Protein Data Bank, the international repository of macromolecular structures (www.pdb.org), currently contains a total of 102 720 structures, accounting for around 40 000 unique proteins. While this is a very large number, scientists have really only scratched the surface of all the potential proteins out there. In addition proteins are often modified, bind cofactors, drugs or inhibitors, and are produced as different splice variants. Thus we can expect generations of future structural biologists to continue in the quest of magnifying molecules to make them accessible and understandable to everyone. **Q**

Wolf-Dieter Schubert is a professor in the Department of Biochemistry, University of Pretoria, specialising in molecular aspects of infectious diseases using Structural Biological techniques. His undergraduate and postgraduate degrees in Chemistry and Crystallography are from the University of Cape Town, his PhD from the Free University of Berlin.