


Crystallography and the atomic anatomy of viruses

David I. Stuart, Jonathan M. Grimes,
Nick Burroughs & Peter P. C. Mertens

 David Baltimore in his 1976 Nobel Lecture observed “*The study of biology is partly an exercise in natural aesthetics. We derive much of our pleasure as biologists from the continuing realization of how economical, elegant and intelligent are the accidents of evolution that have been maintained by selection. A virologist is amongst the luckiest of biologists because he can see into his chosen pet down to the details of all of its molecules. The virologist sees how an extreme parasite functions using just the most fundamental aspects of biological behaviour. ...*”

These words are made flesh by X-ray crystallography. We use relatively soft X-rays (wavelength about 0.1 nm) that shine through our small crystals of virions, illuminating every detail and allowing atomic models to be constructed of the exquisitely ordered symmetrical shells of ‘spherical’ viruses. Knowledge of the three-dimensional positions of the atoms gives a complete static view of the virus and crystallography can, via the somewhat obscure B-factor, indicate molecular flexibility. The results are objective, however, the personal aesthetics of the crystallographer are betrayed in the often riotous choice of colours for their representation. The first virus structure was determined in 1977 by Stephen Harrison and co-workers at Harvard, and over the years a handful of groups around the world have built up a gallery of over two dozen virion structures that cover hosts ranging from bacteria, through plants, to animals. The structures almost exclusively relate to non-enveloped spherical viruses and a challenge for the future is to extend the technique to the less amenable lipid-containing viruses and the beautiful but complex machinery of some bacteriophages. Crystal structures provide snapshots of the virus life cycle which have profound implications for our thoughts on many aspects of their biology. Thus structure has begun to at least partially illuminate questions of virus assembly, evasion of the immune response and cell recognition and entry. Whilst crystallography has led to some unexpected simplifications in our thinking about viruses (for instance, a recurring structural motif suggests that most viruses studied to date may be related), it has also shown the tremendous richness of invention that has been achieved by selecting accidents of evolution.

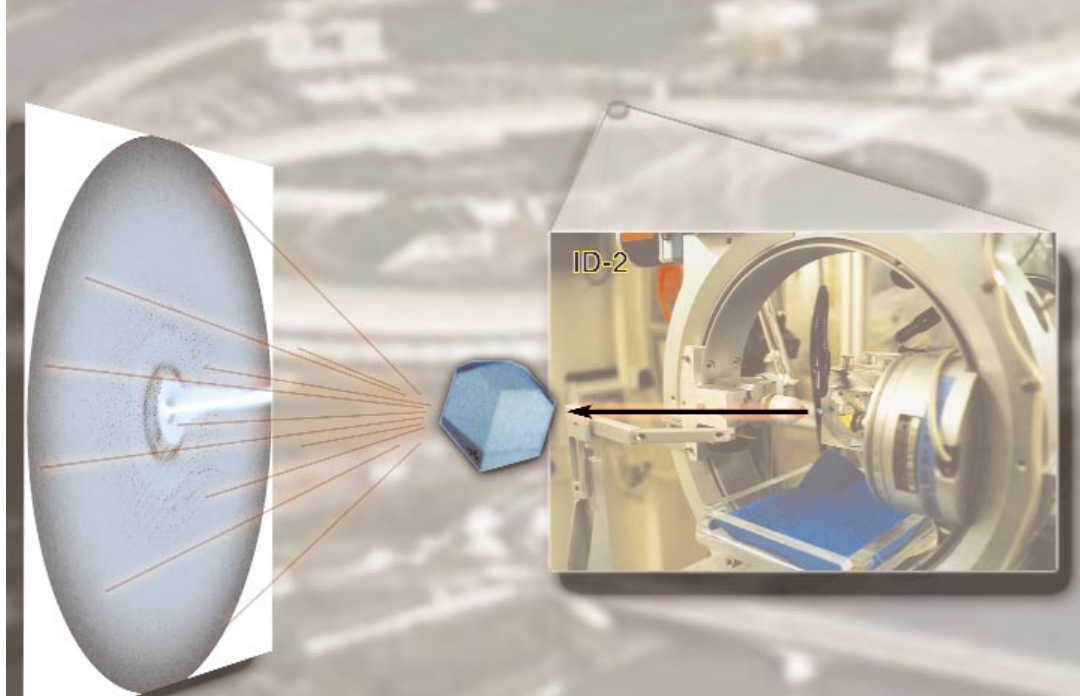
● Small crystals / large viruses

A crystal of an organic compound or small protein contains many trillions of identical molecules whose molecular structure may be visualized by amplifying the feeble scattering of a single molecule. Thus, when illuminated by a nearly parallel beam of monochromatic X-rays, all the molecules in the crystal scatter together in certain directions to give detectable diffracted beams. The concentration of the signal into such diffraction spots, according to rules put forward by Bragg, allows relatively simple compounds to be analysed quite routinely in the lab. Virus crystallography is somewhat more complex but identical in its essentials. It turns out that the necessary first step, that of growing crystals, is often relatively easy for viruses. A solution of pure virus particles, at a concentration of a few milligrams per

millilitre, is simply treated as if it were a small protein and subjected to the battery of crystallization techniques now standardized in commercial kits. Whilst it is perhaps not surprising that robust spherical virions crystallize, it is astonishing that the crystals, held together by only a few flimsy non-covalent interactions, are so beautifully ordered. Naturally, since viruses are large, the diffraction from virus crystals is far weaker than for single proteins. For instance, in the case of the bluetongue virus crystals we have looked at, the number of repeating units across the width of the X-ray beam has been as small as 300, dramatically reducing the amplification of the signal upon which the method rests. For this reason, and because the number of diffracted beams to be measured increases in proportion to the particle size, virus crystallography still presents technical challenges and progress has been quite closely linked to the development of intense X-ray beamlines at synchrotrons. Several third generation synchrotrons are now coming on-line around the world (the first of which was the European Synchrotron Radiation Facility at Grenoble; Fig. 1) which present new opportunities to structural virologists. These machines produce X-ray beams thousands of times stronger than can be obtained with conventional X-ray generators and allow the accurate measurement of very weak data. Furthermore, undulator beamlines are available on these machines which produce very nearly parallel beams, allowing more reflections to be resolved on a detector of a certain size. With the technology now available it is hard to see the limit to what can be achieved. For instance, we have recently solved the structure of the bluetongue virus core that contains almost 1000 polypeptide chains (Fig. 2). Even larger structures could be analysed, provided that they have a well defined structure and provided there are dedicated microbiologists committed to making the many milligrams of pure virus required to analyse such complex systems.

● Problem spots

X-rays cannot be focused sufficiently well to directly image atomic structures. Instead the information in the beams of scattered X-rays is reconstructed with a computer to provide the image. The diffraction pattern obtained is a series of spots which may be thought of as representing the scattering from the virus viewed through a grating. The periodicity of these spots simply reflects the periodicity of the crystal whereas the brightness of the spots reflects the distribution of electrons within the virus. To solve a crystal structure diffraction spots are recorded from a complete set of views of the crystal. Since X-rays quickly damage virus crystals we usually need to add together data obtained from many crystals. Next the spots must be transformed into a map of the electron distribution in the virus. Here we encounter a fundamental problem. The map is constructed from many component waves of different frequencies, amplitudes and directions (any object can be represented in this way, as pointed out by Fourier). Each diffraction spot defines one such component, whose frequency and direction



ABOVE:
Fig. 1. The European Synchrotron Radiation Facility at Grenoble. The beamline, a bluetongue virus crystal and its diffraction pattern are shown in close-up.
PRODUCED BY J. GRIMES & D. STUART

OPPOSITE PAGE:
Fig. 2. The structure of the core of bluetongue virus. The coiled dsRNA genome is shown surrounded by two layers of protein which have been partially stripped away for clarity.
PRODUCED BY J. DIPROSE, J. GRIMES & D. STUART
USING SOFTWARE PREPARED BY R. ESNOUF

is described by the direction of scatter of the diffracted beam and whose amplitude corresponds to the spot brightness. To put the waves together to form the image we need to know their relative phases (this is crucial – in-phase waves augment each other whereas out-of-phase waves interfere!). Unfortunately, when a diffracted beam hits a detector, only its energy is recorded and all phase information is lost. The crystallographer has to find another source of information to reconstruct these missing phases. Fortunately, there are now routine techniques which, for viruses, usually start from rough phase estimates derived from a three-dimensional model. The model might be a crystal structure of a related virus, a lower resolution cryo-EM reconstruction or perhaps a combination of electron microscopy and high resolution X-ray structures of individual protein components.

● Symmetry saves the day

The phases depend on the orientation and position of the model with respect to the axes of the crystal. The icosahedral symmetry of the virus provides us with a key not only to unlock these secrets but also to refine our rough initial estimates of the phases. Icosahedral symmetry consists of a set of 5-, 3- and 2-fold symmetry axes arranged in an absolutely fixed relationship to each other. If we look along a 5-fold symmetry axis, the virus structure will repeat every $360/5 = 72^\circ$ around the axis. The same is true of the scattering from the virus. It is a simple matter, in a computer, to search for such repeats in the measured scattering and thereby lock on to the orientation of the particle. Once correctly orientated the model is moved systematically around the repeating volume of the crystal to find the position that best predicts the observed scattering. Phases may now be calculated, combined with the measured spot amplitudes, and an electron density map calculated. This map will be biased towards our rough model of the structure, adding noise to the image. Again icosahedral symmetry helps us, since the repeating crystal lattice cannot build in the icosahedral 5-fold symmetry. Inescapably, therefore, the virus particle has additional internal symmetry. Since we have located these symmetry axes it is a simple matter to impose icosahedral symmetry to clean up the picture. In practice we use a cyclic procedure, using the cleaned up image to get better phase estimates, which are combined with experimental amplitude measurements to give a new, improved image which is then subjected to further sanitization. Fortunately computers are not easily bored and carry this process through many cycles to produce maps of great clarity.

● Early achievements

The seeming complexity of virus capsids can mask their underlying simplicity. In 1956 Watson and Crick realised that 60 identical subunits could self-assemble to form a closed particle with icosahedral symmetry,

thereby satisfying the biological imperative of genetic efficiency. Some years later Casper and Klug proposed that even more complex assemblies might be made by relaxing the requirement for exact symmetry (giving rise to the term 'quasi-equivalence'). This can be achieved by breaking the icosahedral building block down into a number (T) of sub-triangles, such that the shell is made up of $60T$ chemically identical units. This explains the architecture of many viruses but does not provide a mechanism by which the exact size of such massive assemblies is determined. This is exemplified by tomato bushy stunt virus (TBSV, the first virus solved), where the $T=3$ arrangement of the protein building blocks is exactly as predicted by Casper and Klug but where the architecture is defined by a pathway of controlled conformational switching of the chemically identical subunits during assembly. This combination of quasi-equivalence and conformational switching has now been found in many virus structures. The TBSV structure also revealed a new type of protein fold, an elongated wedge made from long strands of β -structure, often called a jelly roll since its strands are wrapped up as if the chain had been formed in the same way as the American confection of that name. This structure has now become a virtual trademark of viruses, being found in the capsid proteins of an enormous range of viruses (although it is relatively scarce in the proteins of bacteria, plants and animals). The tempting (but unproven) inference is that this reveals unsuspected ancient links between plant and animal, RNA and DNA, and enveloped and non-enveloped viruses. Even if this is true it remains a complete mystery why this structural signature is maintained over such enormous evolutionary distance in very different structural contexts. Very different questions were answered when the first animal viruses were solved in the mid-80s; poliovirus by Jim Hogle and human rhinovirus by Michael Rossmann. These structures immediately led to a clear rationalization of many of the antigenic properties of animal viruses and to an apparent resolution of the conundrum of how viruses maintain binding sites for a cellular receptor whilst evading the immune response. The canyon hypothesis proposes that receptor binding residues are concealed from antibody molecules within a crevice on the viral surface and is still the subject of much discussion.

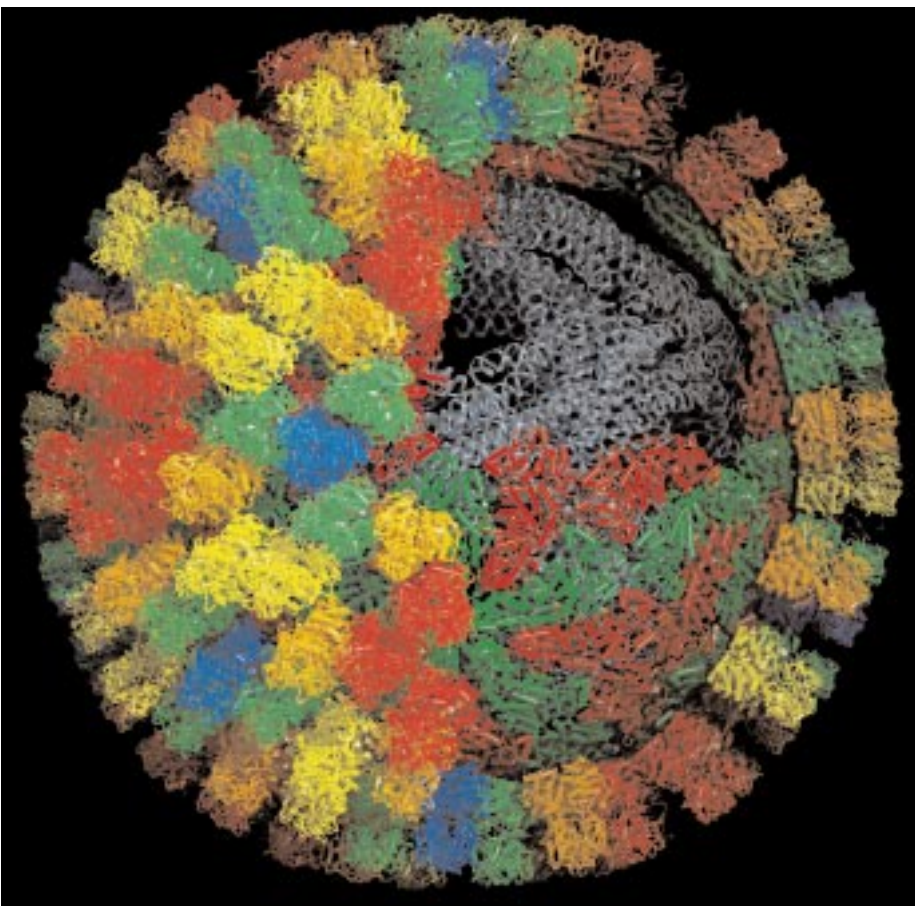
● Moving up a gear: bluetongue virus

Most viruses analysed crystallographically have been around 30 nm in diameter. The publication last year of the structure of the 70 nm bluetongue virus (BTV) core particle (Fig. 2)

demonstrates that much larger assemblies can now be tackled. BTV is a representative dsRNA virus and belongs to the most populous family of such viruses, the *Reoviridae*. The cores of these viruses act as transcriptional machines in infected cells, holding their genome and transcriptional enzymes within a protective shell so that dsRNA is never revealed to the infected cell, eliminating any chance of a cellular response (such as the interferon response) to this unusual nucleic acid. The structure suggests a compelling hypothesis for the assembly of nearly 1000 protein subunits. The core is made from two principal protein components. There is a thin skin covering the genome and enzymes, made from 120 copies of a large protein VP3(T2), which assembles into an icosahedral arrangement via conformational switching to yield a pattern of subunits not seen in other viruses and not predicted by the theory of Casper and Klug. This is clothed in 780 copies of the protein VP7(T13) in an arrangement that follows the theory of quasi-equivalence with greater precision than seen in other virus structures. There is no need for conformational switching in this layer since the VP3(T2) subcore acts as a scaffold upon which the VP7(T13) layer assembles. Intriguingly there is evidence from low resolution EM structures of other dsRNA viruses that the peculiar arrangement of VP3(T2) may be conserved even beyond the family *Reoviridae*. It is tempting to speculate that this reflects some fundamental involvement in the unique biology of these viruses. In most crystallographic analyses of viral capsids there has been little trace of visible structure for the genome. This does not mean that such structure does not exist, merely that the

process of crystallization has laid down viruses in random orientations, blurring the image beyond recognition. The BTV core is very unusual in this respect and we can get a fair idea of how much of the genome (20,000bp), made up of ten unequal segments, is arranged. Unfortunately inappropriate blurring of the image limits the detail and means that interpretations are speculative. Nevertheless, by taking the structure together with information from previous electron microscopical and biochemical analyses we can propose a working model in which the enzymes form transcription complexes under each of the 5-fold apices of the protein shell. These contain three components, which (working outwards from the centre) unwind the genomic dsRNA (a NTP-driven helicase), transcribe a messenger RNA sense copy and cap it before it emerges from the core, ready to initiate protein synthesis. We suspect that the ten RNA segments are neatly coiled around these transcription complexes, so that each segment is set up to run as an independent machine without tangling the enormously long pieces of dsRNA.

Crystallography of viruses has come a long way in the last 20 years. Nevertheless, the structural bases of many of the fundamental biological functions are still unsolved. These functions are often carried out by large multi-component protein complexes. Improvements in the molecular biology and purification techniques for these complexes, along with developments in synchrotron radiation, now provide a way to tackle these difficult problems. We expect crystallography to be in the vanguard of structural approaches to the major functional problems in microbiology.



● Professor David Stuart and Dr Jonathan Grimes are at the Laboratory of Molecular Biophysics, Rex Richards Building, South Parks Road, Oxford OX1 3QU

● Dr Peter Mertens and Mr Nick Burroughs are at the Institute for Animal Health, Pirbright Laboratory, Ash Road, Pirbright GU24 0NF

Further reading

Chiu, W., Burnett, R.M. & Garcia, R.L. (editors) (1997). *Structural Biology of Viruses*. Oxford: Oxford University Press.

Gouet, P. & others (1999). The highly ordered double stranded RNA genome of bluetongue virus revealed by crystallography. *Cell* (in press).

Grimes, J.M. & others (1998). The atomic structure of the bluetongue virus core. *Nature* 395, 470–478.

Harrison, S.C., Skehel, J.J. & Wiley, D.C. (1996). Virus structure. In *Fields Virology*. Edited by B.N. Fields, D.M. Knipe & P.M. Howley. New York: Lippincott-Raven.

Rux, J.J. & Burnett, R.M. (1998). Spherical viruses. *Curr Opin Struct Biol* 8, 142–149.