

UC Irvine

UC Irvine Previously Published Works

Title

A guide to the crystallographic analysis of icosahedral viruses

Permalink

<https://escholarship.org/uc/item/64w5j3pz>

Journal

Crystallography Reviews, 21(1-2)

ISSN

0889-311X

Authors

McPherson, A
Larson, SB

Publication Date

2015

DOI

10.1080/0889311X.2014.963572

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at

<https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

REVIEW

A guide to the crystallographic analysis of icosahedral viruses

Alexander McPherson* and Steven B. Larson

Department of Molecular Biology and Biochemistry, University of California, Irvine, CA, USA

(Received 13 August 2014; accepted 5 September 2014)

Determining the structure of an icosahedral virus crystal by X-ray diffraction follows very much the same course as conventional protein crystallography. The major differences arise from the relatively large sizes of the particles, which significantly affect the data collection process, data processing and management, and later, the refinement of a model. Most of the other differences are due to the high $5\ 3\ 2$ point group symmetry of icosahedral viruses. This alters dramatically the means by which initial phases are obtained by molecular substitution, extended to higher resolution by electron density averaging and density modification, and the refinement of the structure in the light of high non-crystallographic symmetry. In this review, we attempt to lead the investigator through the various steps involved in solving the structure of a virus crystal. These steps include the purification of viruses, their crystallization, the recording of X-ray diffraction data, and its reduction to structure amplitudes. It further addresses the problems attending phase determination and ultimately the refinement of a model. Finally, we describe the unique properties of virus crystals and the factors that influence their physical and diffraction properties.

Keywords: virus crystallography; icosahedral symmetry; non-crystallographic symmetry; electron density averaging; crystallization; X-ray diffraction

	Contents	PAGE
1.	Introduction	4
2.	Production of viruses for X-ray diffraction	5
3.	Purification of viruses for crystallization	7
4.	Crystallization	8
5.	Crystal considerations	12
6.	AFM analysis of virus crystals	14
7.	Preliminary X-ray analysis of virus crystals	20
8.	X-ray data collection	24
9.	X-ray data processing	27
10.	Measures of X-ray data quality	29

*Corresponding author. Email: amcphers@uci.edu

11. Determining the orientation of virus particles in the unit cell	33
12. Probe models	34
13. Heavy atoms and molecular replacement	35
14. A general overview of structure determination	37
15. Refinement of virus crystal structures	39
Notes on contributors	46
References	47
Subject index	54

1. Introduction

The only intact viruses that have been, and probably can be, crystallized and studied by high-resolution X-ray diffraction analysis are relatively small icosahedral viruses. Filamentous or helical, rod-shaped viruses, and others having asymmetric shapes or extreme aspect ratios may be studied by other kinds of X-ray diffraction, such as small-angle scattering or fibre diffraction, but those approaches, in general, do not yield models that are precise at the molecular level. This review will focus on small viruses with icosahedral capsids (sometimes referred to as spherical viruses) like those exemplified in Figure 1. The methodologies and strategies for crystallization, structure determination, refinement, and analysis described here, however, have general applicability to any crystalline virus.

As illustrated for one of the simplest of these viruses, shown in Figure 2, the genomic nucleic acid of the virus is enclosed (or according to accepted terminology, encapsidated) inside a multi-protein subunit container or capsid [1–6] (for a detailed discussion of virus architecture, see the article by DLD Caspar in preparation for a future issue of *Crystallography Reviews*). The nucleic acid may, in principle, be either single- or double-stranded RNA or DNA. In most of the examples studied by X-ray crystallography until now, the nucleic acid has been single-stranded RNA (ss-RNA), though there are notable exceptions [7] such as bacteriophages HK-97 [8] and

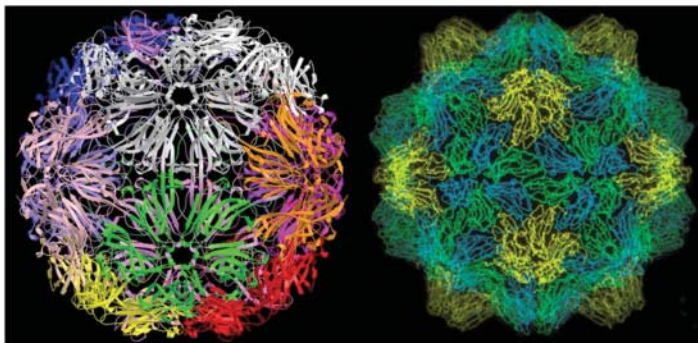


Figure 1. On the left is a ribbon image of the STMV where protein subunits surrounding each of the 12 pentameric vertices are coloured differently so that the symmetry elements of the particle are more pronounced. STMV has a diameter of about 18 nm. On the right is a backbone image of TYMV, a $T = 3$ icosahedral virus. The pentameric capsid proteins are in yellow and the two different conformations of the pseudo-hexameric capsid protein subunits are in blue and green. The diameter of TYMV is about 30 nm.

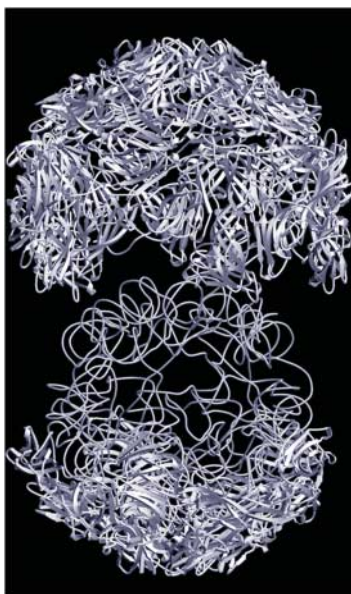


Figure 2. The simplest icosahedral virus, like STMV shown here, comprised of a protein capsid having $T \times 60$ subunits; here $T = 1$. The virion has been separated along a fivefold axis so that the capsid (shown in ribbon representation) is divided into two parts exposing the RNA (shown in tubular form) within. Virus capsids are organized to exhibit $5\ 3\ 2$ point group symmetry. Inside the capsid are one or more strands of either RNA, usually single-stranded, or DNA that code for the capsid protein amino acid sequence, and for some other enzymes that might be required for replication. The peculiar feature of viruses is that they cannot replicate except with the aid of host cell enzymes.

PRD.[9] Host cells, which serve as the sources for the viruses, are highly varied and may come from the plant, animal, or microbial kingdoms.

An excellent general source book for macromolecular structure determination is volume F of the International Tables for X-ray Crystallography.[10] Although not strictly dedicated to virus crystallography, so many of the methods, instruments, and strategies are the same that it also serves as an invaluable guide if the objective is a virus crystal structure. Specific areas of virus structure analysis using X-ray methods are found in other articles in preparation for this and following volumes of *Crystallography Reviews*. General treatments for the X-ray analysis of macromolecules are also easily accessible.[11–14]

2. Production of viruses for X-ray diffraction

Viruses intended for single-crystal X-ray diffraction analysis are produced primarily in plants, in living insects, in cultured eukaryotic cells, especially insect cells, and in microorganisms. In plants, virus production is relatively straightforward, the required techniques are fairly simple, and the level of production generally high. In the latter cases, some significant difficulties may accompany the enterprise, more sophisticated methodology is required, and the yields are usually lower. Thus, it is not surprising that the earliest virus crystallography (see the review by MG Rossmann in this volume [15] and the review by DLD Caspar in preparation for *Crystallography Reviews*) was applied to plant viruses, and that the greater part of the virus structures that have been determined are plant viruses, trailed by insect viruses.

If a range of plant hosts is available for a specific plant virus, the chosen host is often that which provides the greatest amount of tissue, such as cabbage or tobacco, though in some cases

a particular host may otherwise produce exceptional yields. Obviously, virus resistant strains, which have been widely developed and are popular in agriculture, must be avoided. Essentially the same principle applies to other kinds of viruses as well. In infected plants, viruses may be systemic or they may form only local lesions. Satellite tobacco mosaic virus (STMV), turnip yellow mosaic virus (TYMV), and brome(grass) mosaic virus (BMV), for example, infect systemically. One needs only infect a plant on a single stem or leaf and the entire plant becomes infected as the virus passes from cell to cell throughout the tissue.[16,17] Satellite tobacco necrosis virus (STNV), on the other hand, does not become systemic and concentrations of the virus appear only in small localized regions around points on the plant where it was inoculated. Clearly, in terms of production and yield, systemic viruses are preferred.

Plant viruses may have specific vectors in nature, sometimes animals like ruminants that disrupt and wound plants, or more commonly insects. It is usually not necessary to rely on the natural vectors to achieve infection.[18–20] Virus in a buffer or phosphate-buffered saline may simply be mixed with a mild abrasive such as silica or aluminum oxide and gently rubbed on parts of the plant. At the abrasions cells are wounded and the virus enters the tissue and spreads. Host plants, both before and after infection must, of course, be protected from exposure to other plant viruses that may be in the environment to prevent a mixed infection. It is frequently necessary, in order to maximize virus yield, to maintain strict light and temperature regimens, and conscientious watering. In the best of cases, yields of virus on the order of 25 mg may be isolated from 100 g of fresh tissue. Several kilograms of infected plant tissue may be harvested from a greenhouse and, without further treatment, preserved in the frozen state for years without significant loss.

Insect cells are readily infected with viruses, and they are among the easiest to culture and maintain. From the animal kingdom, they have provided rich sources of viruses including, for example, *Nudaurelia capensis* ω virus,[21] Black beetle virus,[22] and Cricket paralysis virus.[23] In some cases, where cells of an insect are difficult to culture, it may be advantageous or efficient to grow large quantities of the insects themselves, infect the populations, and isolate the virus from the insect tissue. Iridoviruses, for example, can be raised in jars containing pillbox bugs [24] to high density. The insects, following death produced by the virus, are simply allowed to decay and putrefy. The virions, however, are extremely stable and when the insect tissue has virtually disappeared, the virus persists.

Viruses produced in other animal cells, principally animal cells in culture, are the most demanding because cell culture is a complicated enterprise that requires a considerable degree of expertise, patience, care, and specialized equipment. Cell culture, however, is the only recourse when the natural viral host is a higher organism. This presupposes that the susceptible cells of that organism can be cultured at all, and this is frequently a tenuous proposition. In addition, because eukaryotic cells in culture do not achieve high densities, the yield of virus per litre of media is low. There are, nonetheless, many examples of viruses so isolated; these include blue tongue virus,[25] rotavirus,[26] adenovirus,[27] and foot and mouth disease virus.[28]

Another means of producing virus is provided by microorganisms. There are two ways of employing bacteria, yeasts, fungi, and other microorganisms as viral sources. The first is to produce a virus naturally by culturing the infected host microorganism. Bacteriophages having spherical capsids and no tails (tails = crystallization problems), or mutants of bacteriophages lacking tails, can be raised to high densities in bacterial cultures and isolated from the media and/or from the cells. The same is true of viruses from yeasts, also amenable to culture, and fungi.

A second approach is to clone the coat protein, or proteins into bacteria, or introduce the genetic material on a plasmid. If successful, when expressed in the bacteria, the viral proteins will self-assemble into particles that generally closely resemble, or are possibly identical to the natural

viral capsid. These are commonly called ‘virus-like particles’ or VLPs. VLPs have, for example, been generated for the capsid proteins of STNV,[29] human immunodeficiency virus,[30] Mason-Pfizer monkey virus,[31] BMV,[32,33] alfalfa mosaic virus (AMV),[34] cowpea mottle virus (CPMV),[35] Ty3 retrotransposon,[36,37] and a host of other viruses.

The use of VLPs opens the door to investigation by X-ray diffraction of many viruses that otherwise would not be available. They simply could not be produced in sufficient quantity. In addition, viruses that are too large to be addressed by single-crystal X-ray diffraction can yield to this approach. This is because the coat protein of the large virus may reassemble, when expressed in bacteria, into much smaller, $T = 1$ or $T = 3$, particles. These may then be crystallized and studied by X-ray crystallography. A persistent question, however, is whether the capsid structure of a VLP is really identical to that of the native virion, or whether it differs in significant details.

3. Purification of viruses for crystallization

A great many, if not most, proteins subjected to X-ray crystallography today are produced by recombinant DNA techniques, that is, after cloning and expression in a bacteria such as *Escherichia coli*. As a consequence of purification ‘tags’ that accompany the transcripts (e.g. histidine tags and maltose binding protein), these proteins can usually be obtained in high purity by relatively efficient and convenient procedures. With the exception of VLPs, viruses offer no such opportunities since they can generally reproduce only in natural host cells.

Viruses, on the other hand, are, so far as biological entities go, rather easy to purify,[38] and indeed, virus purification kits, though usually designed for small scale, are readily accessible from commercial sources. This is chiefly a consequence of the large size of virus particles and relatively high density arising from their nucleic acid component. They are usually larger than any other soluble assembly in the living cell, including ribosomes, but smaller than organelles such as nuclei, lysosomes, microsomes, mitochondria, and other membrane-bounded vesicles. Simple viruses are also usually robust and impervious to physical and chemical assaults that might destroy less stable entities.[18,19]

Most viruses exhibit solubilities in solutions of high ionic strength that are not unlike most soluble proteins.[39,40] Thus, their purification can include salt fractionation. More importantly, however, their size gives them two properties that are extremely useful. First, viruses can usually be isolated from an extract by high-speed centrifugation, as was the first purified virus tobacco mosaic virus (TMV) by Stanley.[41] For high levels of purity, as might be required for crystallization, they can further be banded on CsCl gradients. This also is often useful when there are multiple forms of a virus, for example, the native virions and empty capsids of TYMV that naturally occur in infected tissue.[42] AMV virions also exist in multiple icosahedral forms with different shapes and triangulation numbers depending on the RNA molecules that they encapsidate, and these can be purified from one another by centrifugation on CsCl.[43]

Second, again principally because of size, viruses are susceptible to precipitation using polyethylene glycols (PEGs) of a variety of molecular weights.[44] This allows viruses to be separated from, for example, ribosomes, or from one another. Satellite viruses can be separated from their larger helper viruses using PEG fractionation. Panicum mosaic virus (PMV), for example, precipitates from crude plant aqueous extracts when 4% w/v PEG 8000 is added, while 8% w/v of the same polymer is required to precipitate satellite panicum mosaic virus (SPMV). Following this procedure, both can be crystallized by conventional means.[45,46]

Fragile viruses from less robust hosts obviously require more cautious handling and cannot necessarily be subjected to the rigours of high-speed centrifugation. Nevertheless, purification

is not nearly the problem it might be, and it is not usually an obstacle to crystallography. Collaboration with trained and experienced virologists is, nonetheless, a wise approach.

There is, in isolating and purifying viruses, a necessity for caution and attention to the physiology of the virus. Particles are, for example, frequently sensitive to pH and their physical size and biochemical properties can be gravely altered as a consequence of changes.[47] At elevated pH, or in some cases just in passing from acid to neutral pH, they can swell by as much as 10–15% in diameter.[48–51] Presumably, their exterior surfaces are radically altered in the process as well. BMV, for example, yielded several crystal forms, or none at all, as the pH was increased from an initial 4.5 to 8.0.[32,52] PMV similarly was only crystallized when the pH was maintained below 5.5 throughout the purification process.[53] The changes in the virus as a consequence of pH are generally not readily reversible.

Many viruses have ions bound at specific sites on their surfaces, and the sites are multiplied by the symmetry of the particle. Ions are often divalent cations such as Mg^{++} or Ca^{++} ,[4,47,54] but also anions in some cases.[55] Removal or addition of ions, therefore, can also alter the properties of the capsids and, therefore, their propensity to crystallize, or the quality of what crystals do form.

The degree of purity of a virus preparation is also relatively easy to ascertain. This is so because a virus is composed of both protein, which absorbs ultra violet light most strongly at 280 nm, and nucleic acid that has its absorbance maximum at 260 nm. Thus, every virus has a 260/280 absorption ratio that is characteristic. Because nucleic acid is less often a contaminant that must be eliminated, the higher the 260/280 ratio observed the better. Generally for a pure plant virus such as TYMV, BMV, CPMV, or STMV the ratio is roughly 7:1 to 8:1.[47]

It should be noted in passing that all virus particles in a preparation, even when ‘pure’, may not be precisely identical, though this may have few consequences for its crystallization. This is because identical protein capsids (which determine crystallization) from a single virus preparation may contain different RNA or DNA molecules. Some viral genomes are multi-partite.[56,57] Capsids may contain one segment of a genome, while other capsids contain another.[54] So far as we currently know, however, this internal difference in content is not reflected in the external shell, but we cannot be certain that this will always be true.

4. Crystallization

The crystallization of biological macromolecules of all kinds, proteins, nucleic acids, assemblies, and complexes has been addressed and reviewed extensively.[58–64] No attempt will be made to repeat that here as it is unnecessary. The question arises, however, as to whether there are any obvious or even suspicious differences between the crystallization of viruses and other biological macromolecules. Indeed, one can find such differences only with difficulty. For the most part the crystallization of viruses and the approaches to accomplishing it are no different than for other macromolecules and their complexes. Indeed, the appearance and mechanical properties of virus crystals, like those shown in Figure 3, give no suggestion that they differ in any respects from more conventional protein crystals.

A review of successful virus crystallization conditions (Table 1) would show that the same variety of precipitating agents (ammonium sulphate, sodium phosphate, PEGs, methylpentanediol (MPD), etc.) were used with viruses as with most crystalline proteins (Protein Data Bank (PDB) [65]). The pH of successful crystallization experiments generally tends towards the acid side of neutrality, likely reflecting the greater stability of most viruses there, and the tendency to swell at higher pH. Use of detergents, even non-ionic detergents, is essentially absent and specific additives [61,66] are rare. In addition, the most common methods for producing supersaturation, hanging and sitting drop vapour diffusion, microdialysis, batch, and free interface

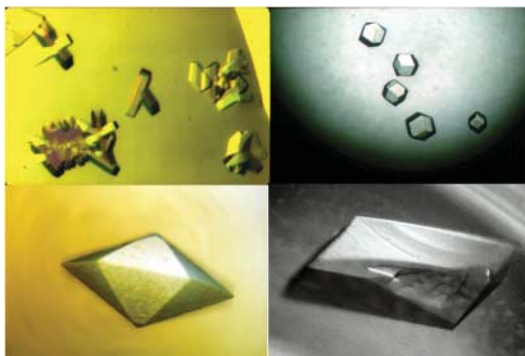


Figure 3. Shown here are hexagonal plates of CMV (upper left), cubic crystals of DYMV (upper right), a hexagonal crystal of TYMV (lower left), and a monoclinic crystal of BMV (lower right).

Table 1. Crystallization precipitants for viruses solved by X-ray diffraction as reported in the PDB.

Precipitant	Number of crystals	Precipitant amount
(1) PEGs < 1000	11	20–30% w/v most common
(2) PEGs 1000–8000	62	See Figure 4 for % distribution
(3) PEGs > 8000	None	
(4) Ammonium sulphate	10	1.5–2.5 M most common
(5) Other salts	18	1.0–2.0 M most common
(6) MPD and hexanediol	5	10–50% v/v
(7) Ethanol or propanol	2	15–25% v/v
(8) Numerous virus crystals obtained by adjustment of pH – no reported precipitant		

diffusion [58,63] predominate. The only obvious difference is that the virus concentration in virus crystallization trials is usually lower, in terms of milligrams per millilitre, than for most proteins. Generally 3–5 mg ml⁻¹ is sufficient to produce diffraction size crystals of viruses, whereas concentrations four times that, or more, may be required for proteins.

Examination of crystallization conditions detailed for 110 unique crystals in the PDB shows that the vast majority, about 62%, were obtained with PEGs having molecular weights between 1000 and 8000. In many of the PEG dependent mother liquors, some salts were also included and often in not insignificant concentrations of 0.2–0.3 M. As further shown by Figure 4, PEG concentrations were low, generally between 1% and 5%, in comparison with those used for more conventional protein crystallizations, which tend to be around 10% and higher.[67]

An additional 11 viruses were crystallized using PEGs of molecular weight less than 1000, along with 5 others crystallized from MPD or hexanediol. For these crystallizations, the range of effective PEG (or MPD) concentrations was 14–30% w/v (with a single virus at 50% MPD). Two viruses were crystallized from ethanol (14–17%) and isopropanol (24%). Viruses have been crystallized over a wide range of pH from 3 to 9, although the majority around neutrality. As illustrated by the histogram of Figure 5, however, the distribution clearly favours the low side of pH 7, again likely reflecting the greater stability of many viruses at acidic pH, and their tendency to become less uniform in structure at alkaline values.[47]

Of the virus crystals, 25–30% were grown using some salt as the principal precipitating agent (though when combined as well with some low concentration of PEG, it is difficult to make a distinction). Ammonium sulphate was most commonly used (~ 12%) and over a concentration range of 0.5–2.5 M, with most successes clustered about 2 M. In addition to ammonium sulphate,

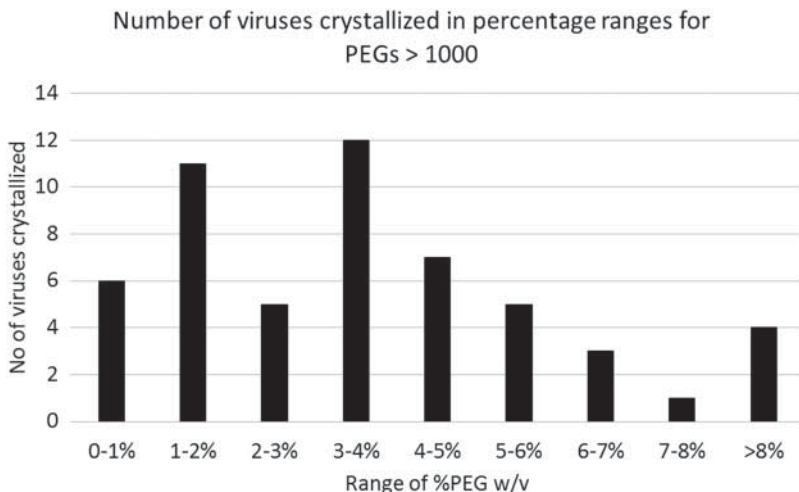


Figure 4. Illustrated is the histogram of the PEG (> 1000) concentrations used for the crystallization of viruses in the PDB.

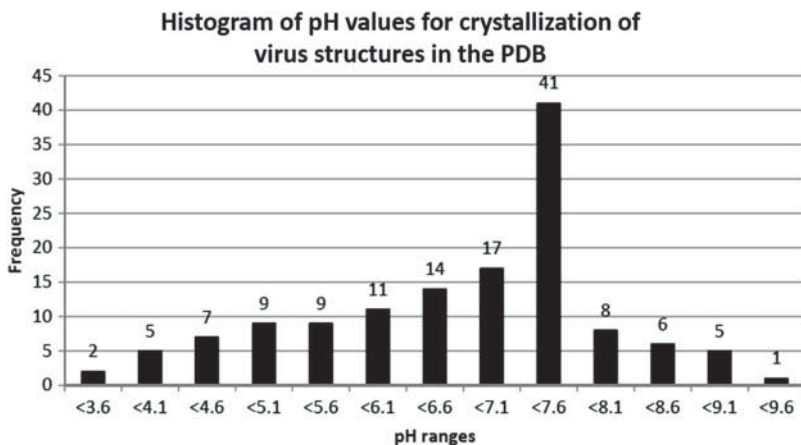


Figure 5. Histogram based on data from the PDB showing the pH at which viruses have been crystallized.

other salts used to promote crystallization were sodium salts of acetate, chloride, and formate; ammonium salts of phosphate, acetate, and formate; and lithium sulphate. These were used in concentrations ranging from 1 to 3.5 M with most successes clustering about 2.5 M. It should also be noted that some virus crystals were obtained from relatively low ionic strength buffers simply by adjusting the pH to a minimum point of virus solubility.

As with some conventional protein molecules, but perhaps more so because of their high symmetry, individual viruses may crystallize in a diverse variety of crystallographic unit cells. This is particularly true as the pH is varied (see, for example, the crystals of BMV in Figure 6), or the precipitant is varied between different salts, or changed from, for example, ammonium sulphate to PEG. The polymorphs may have widely different solvent contents, degrees of order, or vary widely in their diffraction properties. Sesbania mosaic virus, for example, crystallizes in seven distinctly different unit cells according to the PDB, and human rhinovirus crystallizes in at least six forms.

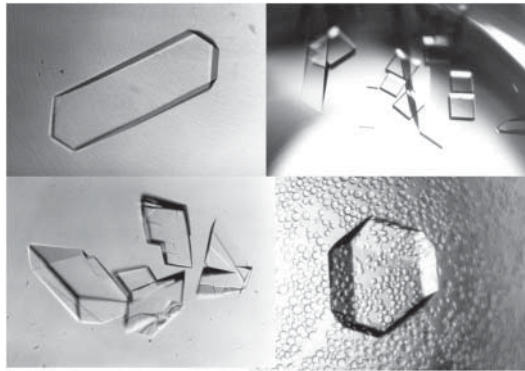


Figure 6. An example of multiple crystal forms of a virus are the four crystals of BMV shown here, each having a unit cell different in symmetry and dimensions from the others. The crystals were grown under similar conditions, but the pH of the mother liquor was varied over the range 5.5 to 7.5.

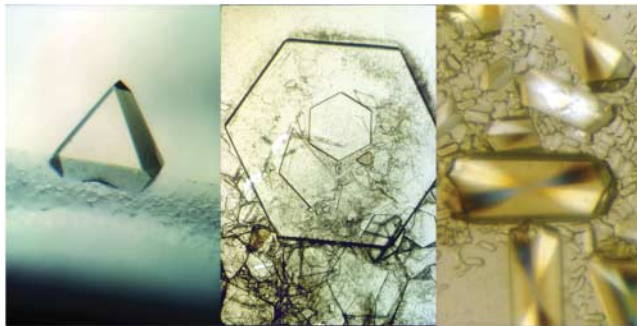


Figure 7. Multiple crystal forms are common with virus crystals, as with most macromolecular crystals, reflecting a lattice maintained by a large array of weak intermolecular interactions. Three crystal forms of STMV are shown. From left to right, the forms are cubic, monoclinic, and orthorhombic. The largest dimension of the crystal in each case is between 1 and 1.5 mm, very large by most macromolecular crystal standards.

There is one particularly challenging requirement in virus crystallization (there may, of course, be more for a specific virus) and that is, the crystals for diffraction analysis must be relatively large in size. Optimally, they would be of the sizes of the STMV crystals seen in Figure 7. While micro-beams, high-intensity synchrotron radiation, cryopreservation, and other technological advances have reduced the necessary dimensions of protein crystals to a few tens of microns, this is not usually true for virus crystals. One must still attain virus crystal dimensions measured in fractions of a millimetre (sometimes large fractions). This requirement stems from the very large unit cell dimensions of virus crystals that may range from 200 Å to over 1000 Å, and the large asymmetric units that will always be some significant fraction of the entire virus, and sometimes the entire particle. Because average intensity of reflections is inversely related to unit cell volume and asymmetric unit size, the crystals must be large just to produce strong enough intensities that can be accurately measured. Crystallization trial volumes of nanolitre volumes, often the standard when robotic methods are used for conventional protein crystallization, are, therefore, seldom of value for viruses except possibly to identify initial conditions.

5. Crystal considerations

For roughly the past 20 years, most X-ray diffraction data have been recorded from protein crystals, and complexes such as ribosomes, that have been flash-cooled in liquid nitrogen and preserved in a cryo-stream during exposure.[68,69] This was both to gain advantage (reduced scaling, less radiation damage [69,70]) and often a practical necessity. Virtually all of those data, particularly the high-resolution data, were collected at synchrotron sources which supplied very high flux density X-ray beams.[71] Crystals simply could not withstand the radiation doses for any useful period of time in the absence of cryo-cooling. Combination of flash-cooling with very high-intensity sources also meant that crystals of decreasingly smaller sizes could be used for data collection, a further advantage since it obviated requirements in the crystal growth phase of a project.

It appears that investigators have frequently encountered difficulties in freezing virus crystals. As discussed above, however, unit cell considerations do not allow for small virus crystals as are now common for most proteins. They just do not produce sufficiently high-intensity X-ray reflections for accurate measurement. Thus, even if virus crystals can be flash-cooled, they still must be large crystals. The problems in freezing virus crystals have been variously attributed to their high solvent content, the large volume of solvent within the particles themselves, and the large interstitial spaces between particles in the lattice. These explanations, though possibly contributors, are probably not sufficient.

Atomic force microscopy (AFM) studies of various icosahedral plant virus crystals (STMV, BMV, TYMV, PMV), *in situ*, during growth, suggest another explanation.[72–75] AFM analyses indicate that virus crystals have a relatively high density of defects and that the defects include the incorporation of large foreign particles, misoriented microcrystals, anomalous virus particles and lattice vacancies. These produce more or less localized disorders, and tolerable disruptions to the lattice, in so far that growth continues. More importantly, they also include large numbers of stacking faults or planar defects. Interestingly, no screw dislocations have yet been observed in virus crystals, though they are common in protein and conventional crystals.[58,61,73,76]

The planar defects, which subdivide the virus crystals into sectors, or domains, are responsible for their mosaic character (see below). The stacking faults also serve as tributaries and reservoirs in the lattice where solvent accumulates and flows. When winter comes and the cracks in the pavement fill with rain or snow that subsequently freezes to ice, the cracks expand and eventually the concrete may shatter. The same process likely exists for macromolecular crystals that contain both large amounts of solvent [58,77–79] and a high density of planar defects. This is the more likely explanation for the problems that arise in cryo-preservation of virus crystals.

Virus crystals are additionally susceptible to damage from freezing because, as discussed above, they must be large in size. The number of defects as well as their extent is proportional to crystal volume. In addition, defects produce both local and long-range perturbation and strain within crystals, and the accumulated lattice strain is also (probably in a nonlinear manner) proportional to volume. The lattice of large crystals always experiences greater stress than that of small crystals.[80,81] The end result is that virus crystals, by virtue of their size, have high defect densities, high solvent content, and an elevated degree of lattice stress baked into them. It is not surprising then that the trauma inflicted by sudden exposure to cryogenic temperatures causes severe disruption, cracking, shattering or, at the least, a significant increase in mosaicity.

In passing it should be noted that the most common technique used in flash-cooling macromolecular crystals [68,69] is to select them from their mother liquor with a small fibre loop, pass the looped crystal rather quickly through a cryo-preservative solution (e.g. 20% glycerol or ethylene glycol) and then plunge them into liquid nitrogen. The common assumption with this procedure is that damage to crystals results primarily from the freezing of solvent (water) about

the surface layers of the crystals. That is, an ice shell forms about crystals that compresses and crushes them. Cryo-preservative solutions are intended to eliminate that shell. Passing a crystal quickly through a cryo-preservative, however, may not allow diffusion of the cryo-protectant into the crystal and the replacement of the water in the defects and vacancies. These may drain or exchange more slowly. Water, upon freezing, will then cause the crystals to crack at domain boundaries.

With virus crystals it may be advantageous to expose them for longer periods to the cryo-preservative before freezing. In addition, anything that can be done to reduce defects, such as enhanced purification, should be undertaken. In spite of the difficulties and many failures, a substantial number of virus crystals have been successfully frozen for X-ray data collection. In most cases, rather complex concoctions of cryo-preservatives have had to be formulated and tested by trial and error. Some examples are STMV,[69,82] BMV [52], PMV,[46] and TYMV.[83] Crystals of smaller, $T = 1$, viruses such as STMV and SPMV have proven easier to flash-cool, and this follows from their smaller unit cell dimensions and the arguments presented above. $T = 3$ viruses have shown themselves to be more challenging, and viruses of even larger sizes and greater T numbers the most difficult of all.

If virus crystals cannot be frozen, and it is well worth extensive efforts to successfully freeze them, then the only recourse is to record X-ray data at room temperature. Conventional X-ray sources that allow longer lifetime in the X-ray beam, and that are commonly sufficient for high-resolution data collection on conventional protein crystals, generally do not provide sufficient intensities for virus crystals of $T > 1$ particles or unit cell dimensions exceeding about 250 Å. Hence, to acquire good X-ray data, it usually is essential to employ synchrotron sources. The trade-off is that synchrotron sources provide measureable reflections, but the crystals, at room temperature, suffer severe radiation damage and deteriorate rapidly.

The investigator must do the best he can under these circumstances. For fairly robust $T = 3$ viruses, we have found that up to about 6 minutes of total exposure time to radiation produced by most second-generation synchrotron sources can be used to obtain still useful data before the crystal is exhausted. This interval may allow 2–4 exposures, usually about 0.5° of rotation each, to be recorded. The last exposure in the set is of course always questionable and has to be evaluated with care, but this concern is lessened somewhat when a large number of exposures are recorded and scaled.

The overall objective is to collect as many acceptable exposures (frames of data) as possible, scale them together, and then eliminate those that contribute more error than information. The saving grace with this approach is that scaling of many exposures collected at room temperature is surprisingly good. Scaling of data clusters or sets collected at cryo-temperatures from multiple crystals on the other hand scale poorly, if they can be scaled at all. Another positive consideration is that virus crystals usually yield an abundance of independent reflections in proportion to the size of the molecule (usually the capsid protein) that must be solved because of the non-crystallographic symmetry (NCS) inherent in the virus particle.

For room temperature data collection, crystals must be mounted in what was once called ‘the conventional manner’ in sealed glass or quartz capillaries.[62,84,85] The art of mounting crystals in this way requires, in addition to ‘grace under pressure’, patience and skill. The art has been nearly lost over time due to the successes of cryo-crystallography, but may be experiencing a revival. In any case, proficiency can be acquired through practice. If it is the only way to obtain the X-ray data, then that is usually sufficient inducement to merit the commitment.

A somewhat simpler method that has more recently appeared is to mount or secure the crystal in a fibre loop, as for cryo-crystallography, but then cover the loop and crystal with an envelope of thin, transparent plastic (Mitigen Co., Ithaca, NY). This can work as well as capillary mounts but due to slow loss of water the arrangement is only useful for data collection for four to six

hours. This, however, is ample time to record several frames of data on a synchrotron source at room temperature. Indeed, because a virus crystal at room temperature can only tolerate a few minutes exposure of such a source, in total, then it may not even be necessary to enclose the crystal in an envelope. One can simply trap a crystal on a loop, align it in the X-ray beam and collect a cluster of frames before significant dehydration occurs. This has been done in a few cases, but it does entail risk.

Room temperature data collection of virus crystals carries an additional implication. Because success depends on the scaling and merging of reflections from many small data sets, it means that many large crystals must be grown. With flash-cooling, all data can conceivably be acquired from a single crystal. Many large crystals of any macromolecule is a daunting, formidable challenge, but one which has been overcome by many investigators. The structure determination of the bacteriophage HK-97, which required 720 separate crystals, is one such heroic example.[8,86]

The discussion above would suggest that virus crystallography presents some unique problems at the crystal growth stage, and this is largely true. There is one distinct advantage that icosahedral viruses have, however, that appreciably ease issues at the crystallization stage, and that is their high symmetry. It has been observed [62,87,88] that symmetry in macromolecules tends to advance the probability of crystallization. Indeed, vast numbers of symmetrical protein oligomers have been crystallized and their structures determined. In a majority of cases, symmetry elements of the oligomer or particle were incorporated entirely, or at least in part, into the ultimate crystallographic symmetry. There is now little argument against the fact that symmetry promotes crystallization.[89]

Icosahedral virus particles [2–6,90] and see article by DLD Caspar in preparation for a future issue of *Crystallography Reviews*, have exact $5\ 3\ 2$ point group symmetry relating their 60 identical asymmetric units (coat protein subunits in the case of $T = 1$ viruses). Although the fivefold symmetry cannot be incorporated into the space group symmetry of a crystal, twofold and threefold symmetry elements can. Thus, icosahedral viruses are often centred on crystallographic special positions, on twofold or threefold axes, and at 23 , 222 or 32 symmetry points.

Icosahedral viruses of triangulation number $T > 1$ possess quasi-symmetry elements as well, such as quasi-sixfold axes.[2,5,6] Quasi-symmetry cannot be incorporated into the space group symmetry, but there is some likelihood that the periodic nature of the quasi-symmetry and the isotropic shape of the overall particle may contribute to favourable and repetitive lattice interactions. The end result is that, in comparison with most other biological macromolecules and assemblies, viruses are fairly easy to crystallize once they have been obtained undamaged and in a pure form.

6. AFM analysis of virus crystals

In addition to earlier studies [91,92] of virus crystallization using quasi-elastic light scattering (QELS), extensive AFM studies have also been carried out on virus crystals *in situ*. [93] Viruses proved to be particularly valuable as samples in QELS investigations, because their large size meant that they produced a strong scattering signal, as did their aggregates. This was especially true in studies of prenucleation and nucleation events. Viruses were equally valuable in AFM studies because their size allowed them to be seen as single particles and their incorporation into crystal lattices directly visualized. Figure 8 provides examples of AFM images of the surfaces of several $T = 1$ and $T = 3$ virus crystals where the individual particles composing the lattice are clearly defined. Figure 9 shows that in some cases, like that of TYMV, individual pentameric and hexameric capsomeres composing the capsids of single particles could be observed. Thus, growth kinetics, defect formation, and other features of crystal growth could be visualized at

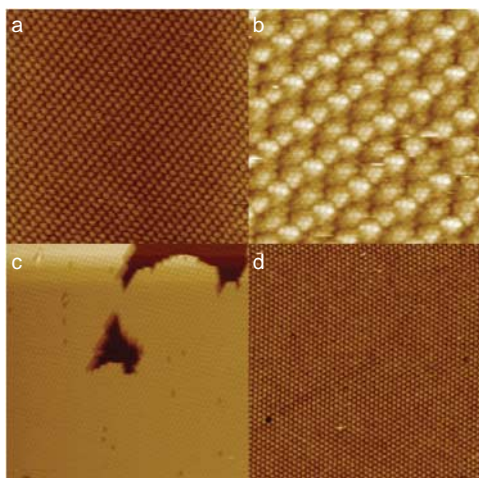


Figure 8. Virus crystals offer a unique opportunity for study by AFM while still growing in their mother liquor. Such studies provide valuable measures of thermodynamic and kinetic parameters that govern growth and form. The advantage of virus crystals is that single particles can be visualized entering or leaving the step edges of the developing surface lattice. Seen here are (a) BMV, (b) TYMV, (c) CMV, and (d) STMV. Scan areas are (a) $500\text{ nm} \times 500\text{ nm}$, (b) $300\text{ nm} \times 300\text{ nm}$, (c) $1.5\text{ }\mu\text{m} \times 1.5\text{ }\mu\text{m}$, and (d) $1.3\text{ }\mu\text{m} \times 1.3\text{ }\mu\text{m}$.

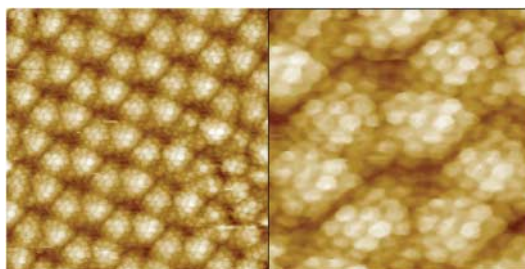


Figure 9. An illustration of the resolution of the AFM technique when applied to virus crystals is shown. On the left is a moderate resolution image of the surface of a growing hexagonal TYMV crystal and on the right is a higher resolution scan showing several particles where capsomeres of individual virions can be discriminated. Scan areas are $300\text{ nm} \times 300\text{ nm}$ and $140\text{ nm} \times 140\text{ nm}$, respectively.

essentially molecular resolution. This is not generally possible with most proteins, though there are some exceptions.[94]

AFM investigations, focused primarily on STMV, BMV, TYMV, cucumber mosaic virus (CMV), and PMV, have been reviewed [2,5,6,74,75,95] and thus need not be repeated here. A few observations from that work are, however, worth recounting. The observations, to some extent, emphasize those features of virus crystals that discriminate them from other macromolecule crystals, and certainly, small organic molecule crystals.

Virus crystals grow from solutions, as do most conventional organic molecule crystals and all other macromolecular crystals, by what is classically referred to as sequential layer addition. Layer addition in the face normal direction relies on the generation of terraces and growth steps by two-dimensional nucleation and/or by spiral dislocations.[80,81,96,97] It pictures the ordered addition of individual molecules at the resulting step edges, by tangential growth, at a rate determined by the level of supersaturation. The only really distinctive difference between virus crystals and protein crystals that has been observed by AFM is that the generation of growth

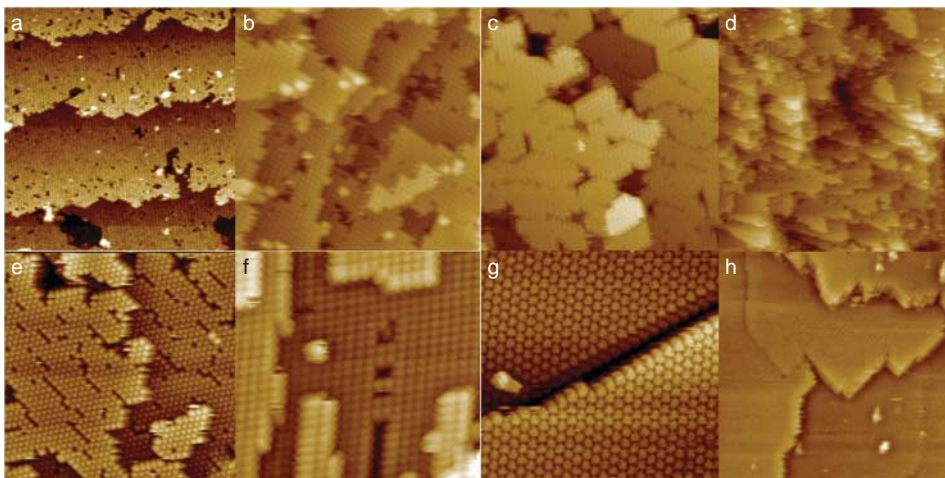


Figure 10. Virus crystals develop, as do almost all crystals grown from solution, by the sequential addition of ordered molecules to their surfaces. The planes expand laterally from an initial island, or two-dimensional nucleus on the surface, by incorporation of virus particles to the edges of the expanding planes, the so-called step edges.[80,81] Here step edges are seen on a variety of growing virus crystals. In (a), (c), and (e) are STMV crystals, (b) is BMV, (d) TYMV, (f) PMV, and (g) and (h) CMV. Scan areas are (a) $2\ \mu\text{m} \times 2\ \mu\text{m}$; (b) $2\ \mu\text{m} \times 2\ \mu\text{m}$; (c) $1\ \mu\text{m} \times 1\ \mu\text{m}$; (d) $7\ \mu\text{m} \times 7\ \mu\text{m}$; (e), (f), and (g) $700\ \text{nm} \times 700\ \text{nm}$; and (h) $2.5\ \mu\text{m} \times 2.5\ \mu\text{m}$.

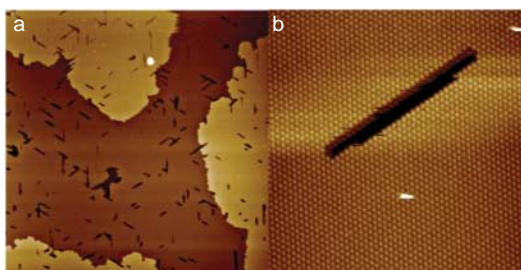


Figure 11. Among the many unusual impurities that are incorporated into virus crystals, and other macromolecular crystals as well, are the remnants of microbes that have degraded in the mother liquor. Here is a fairly low magnification image in (a) and a high magnification image in (b) showing the scars that remain upon the incorporation of what are presumably cytoskeletal fibres from dead microbes into the crystallographic lattice. The phenomenon illustrates not only the variety of impurities that might affect crystal quality, but also the extremely forgiving nature of the virus crystal lattice. Scan areas are (a) $25\ \mu\text{m} \times 25\ \mu\text{m}$ and (b) $1\ \mu\text{m} \times 1\ \mu\text{m}$.

steps by spiral dislocation does not appear to be a growth mechanism for virus crystals. Face normal growth seems to be exclusively due to two- and three-dimensional nucleation on existing surfaces.[58,98]

Virus crystals, because of the large growth step heights at the advancing edges (Figure 10), incorporate vast amounts of impurities into their lattices. That is, the growth steps, as they move across the surfaces of crystals, singly or in step bunches,[99] sweep everything before them, like great waves, into the channels and interstices between particles. This has been seen to include fibres of various sorts (Figure 11),[100] dust particles,[101] misoriented microcrystals (Figure 12),[102,103] and as seen in Figure 13, anomalous and mutant particles.

A particularly striking case was recorded for the lattice of a monoclinic crystal of PMV, a $T = 3$ plant virus of 30 nm diameter. Preparation of this virus required separating PMV from its satellite

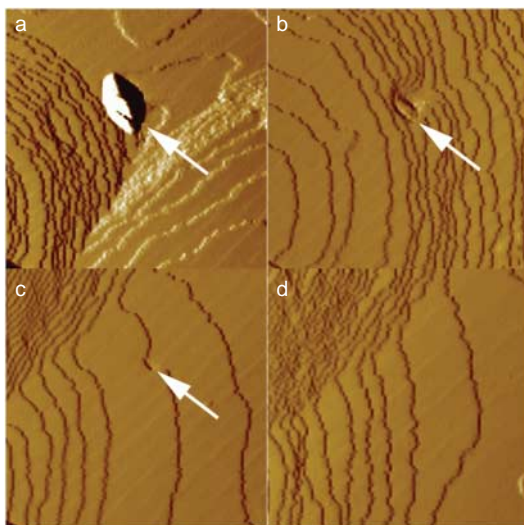


Figure 12. Microcrystals that form in the mother liquor of a growing virus crystal can sediment on the developing surfaces of larger crystals and be incorporated intact. Thus, a large virus crystal may have embedded within it microcrystals having random orientations with respect to the greater crystal lattice. This is illustrated by a sequence of AFM images of a growing STMV crystal made at 15 min intervals showing a sedimented microcrystal (indicated by an arrow) being inundated and submerged by the moving step edges on the surface of the larger crystal. The scan areas are $10.5 \mu\text{m} \times 10.5 \mu\text{m}$.

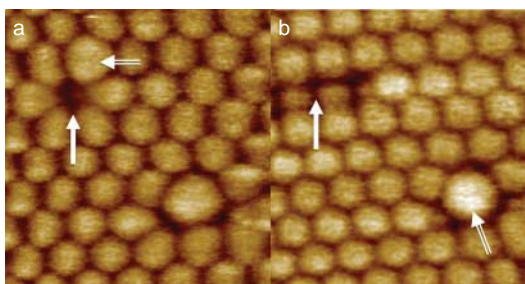


Figure 13. In (a) is the surface layer of virus particles of a BMV crystal and in (b) that of a CMV crystal. In (a) an arrow marks a point defect, or vacancy, and the double arrow identifies an anomalously large, aberrant particle that has been wholly incorporated into the crystal lattice. In (b) the single arrow indicates a line of missing particles, and the double arrow again the site of incorporation of an oversized virion. Virus preparations contain many aberrant particles, many of which are nonetheless incorporated into the crystal lattice. Note, however, that when this occurs there is local disorder in the surrounding lattice. The scan areas are $200 \text{ nm} \times 200 \text{ nm}$.

SPMV ($T = 1$, 17 nm diameter) by PEG fractionation. As a consequence, SPMV remained a prominent contaminant in PMV crystallization mother liquors. In Figure 14, virions of SPMV can be seen incorporated into the crystal lattice of PMV in the interstitial spaces between PMV virions.[53,104]

Another interesting case is typified by BMV.[32] In Figure 13, the lattice of a BMV crystal is seen to contain not only normal 30 nm diameter BMV virions, but occasionally, distinctly larger, anomalous virus particles of greater diameter. The lattice of an orthorhombic STMV crystal in Figure 15 illustrates another common phenomenon in which there are frequent absences in the lattice (called vacancies) of single particles, clusters, and lines of particles in the lattice. The images demonstrate that in STMV crystals, as in most other macromolecule crystals, some unit

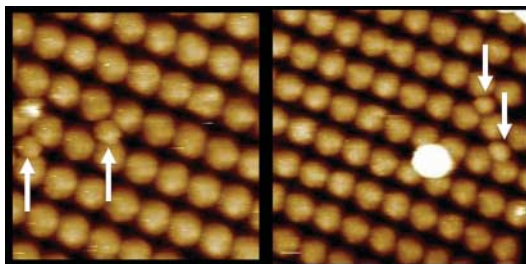


Figure 14. In the left panel, the arrows indicate 17 nm diameter SPMV virions residing in the space created by four 30 nm diameter PMV particles in the PMV crystal lattice. In the right panel, two SPMV particles, indicated by arrows, occupy such spaces. The large white object in the right panel is a PMV virion moving on the surface of the crystal that has not yet found its place in the lattice. Scan areas are 250 nm \times 250 nm and 300 nm \times 300 nm, respectively.

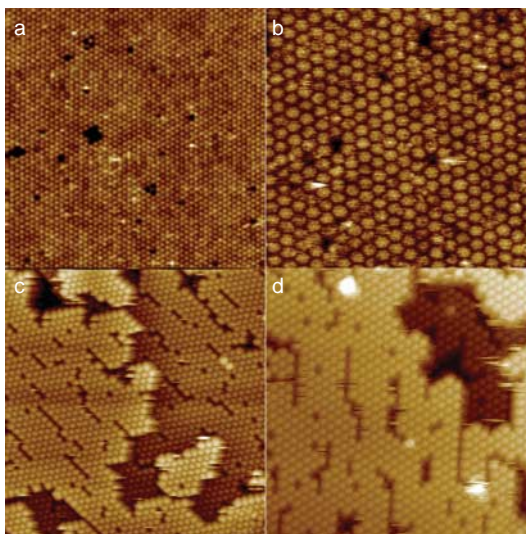


Figure 15. In (a) and (b) are images of the surfaces of growing BMV crystals, and in (c) and (d) growing STMV crystals. In the surface layers of the crystals many point defects are present due to vacancies, or the absence of particles. Particularly in (c) and (d) line defects created by a linear series of missing virus particles are present. Neither the point defects nor the line defects are filled before the next, superior layer is completed. Thus the defects are present in the final crystal just as seen here. Scan areas are (a) 1.5 μ m \times 1.5 μ m, (b) 542 nm \times 542 nm, (c) 800 nm \times 800 nm, and (d) 650 nm \times 650 nm.

cells, perhaps as many as several per cent, remain unoccupied. In spite of these defects, the crystals diffract to unusually high resolution.

In Figure 12, also AFM images of an STMV crystal, it can be seen that during the course of its growth, microcrystals, presumably having formed spontaneously in the mother liquor, sediment on the surfaces of a larger, growing crystal. These too are incorporated, misoriented as they are, into the larger crystal. Thus, we see that virus crystals may be inordinately permeated with a wide array of different impurities that exceed, probably by several orders of magnitude, the quantity consumed by conventional crystals, and even protein crystals.[102]

It might be expected that the extensive impurity incorporation observed by AFM would gravely interfere with crystal growth and even cause it to cease. It would certainly do so for conventional crystals. It might seem extraordinary, in fact, that large virus crystals can even be obtained. They do, nonetheless, grow to large dimensions because the lattices of virus crystals

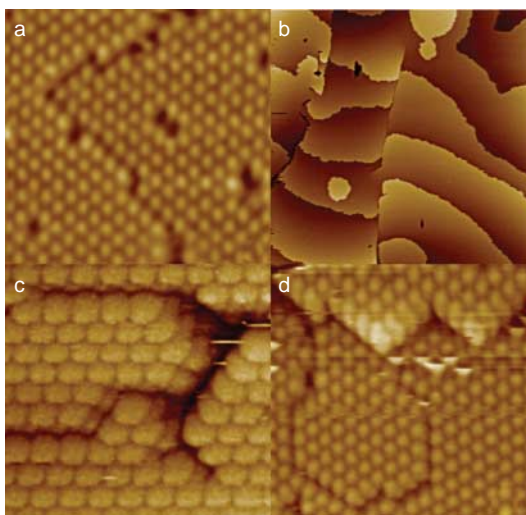


Figure 16. In some instances, step edges originating from different two-dimensional nuclei on developing crystal surfaces do not merge seamlessly because of some displacement of the edges. Thus, different sectors of a single crystal may be slightly displaced with respect to one another, thereby creating domains or mosaic blocks. With virus crystals, the details of these domain boundaries, also known as stacking faults or planar defects, are clearly visible. Here are seen domain boundaries in (a) and (d) crystals of CMV, in (b) STMV, and (c) PMV. Scan areas are (a) $1\ \mu\text{m} \times 1\ \mu\text{m}$, (b) $40\ \mu\text{m} \times 40\ \mu\text{m}$, (c) $350\ \text{nm} \times 350\ \text{nm}$, and (d) $1\ \mu\text{m} \times 1\ \mu\text{m}$.

appear to be unusually forgiving. They can absorb extensive insults and offenses and simply grow around them. Defects (see below) are created as consequences of impurity incorporation, but these too fail to prevent further growth from proceeding. Apparently this is due to the plastic nature of virus crystals, likely a consequence of particle elasticity and size, the high solvent content of the crystals, and the large spaces between particles in the lattice.

As noted above and illustrated by STMV and BMV (Figure 15), lattices can exhibit point defects and vacancies, and even line defects due to strings of vacancies. These defects are relatively innocuous, localized, and while the absent lattice points fail to contribute to the Bragg scattering, any damage that is consequential to diffraction is limited. Similarly, incorporation of anomalous particles results in some local disorder, as seen in Figure 13 for BMV, but it too is fairly restricted and the effects are not serious. Because no spiral dislocations are apparently present in virus crystals, likely because of the large step heights, there are no long distance line defects passing through crystals along screw dislocation axes.

What is found in relative abundance in virus crystals are stacking faults and planar defects of various kinds. These arise when separate growth terraces and planes (from two- and three-dimensional islands) encounter one another on the developing layer and their step edges fail to merge and knit in a flawless manner, as seen in Figure 16. That is, there is some vertical displacement of a fraction of the step edge height between apposing steps. When this occurs, then uniform forward advancement of step edges (Figure 10) is disrupted and redirected as seen on the surface of an STMV crystal in Figure 17. A consequence of this is that vertical dislocations are created between the molecules and unit cells comprising one expanding plane and those of others. The net effect is to effectively create mosaic blocks within the crystal (Figure 18) and produce a spread in the Bragg angles for reflections, and therefore the width of observed intensities.[71,105] In some virus crystals, such as STMV [72] or CMV,[106] the planar defects are very common and the crystals exhibit what we call a high defect density (Figure 18), orders of magnitude greater than for conventional crystals. This elevated defect density is likely a major

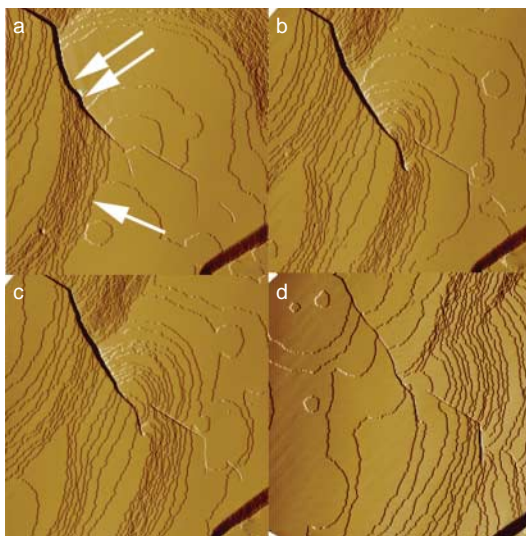


Figure 17. In four successive AFM scans of the developing surface of an STMV crystal, the pattern of step edge movement in the neighbourhood of a planar defect is visualized. The moving steps (indicated by a single arrow) cannot cross the domain boundary established by the stacking fault (indicated by the double arrows) but are forced to flow around its lower end in order to fill the opposite side. The defect, then, affects not just one layer, but is propagated throughout the crystal to create a planar barrier between blocks. The time points are (a) $t = 0$, (b) 420 s, (c) 600 s, and (d) 1560 s. The scan areas are $22 \mu\text{m} \times 22 \mu\text{m}$.

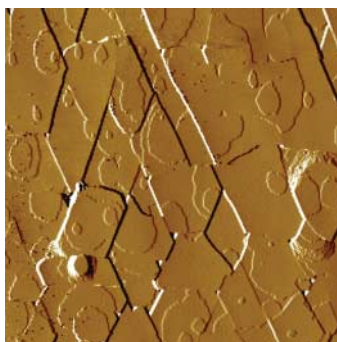


Figure 18. Domains or mosaic blocks are evident from the pattern of planar defects present on the surface of this crystal of STMV. The scan area is $80 \mu\text{m} \times 80 \mu\text{m}$.

constraint on the resolution of the diffraction patterns yielded by virus crystals, which, it might be noted in passing, varies for the diversity of virus crystals over a wide range (Figure 19).

7. Preliminary X-ray analysis of virus crystals

The symmetry properties of icosahedral viruses have been dealt with extensively in the literature [2,3,5,6,107] and will be further discussed by DLD Caspar in a review in preparation for a future issue of *Crystallography Reviews*, so no attempt will be made to comprehensively review that here. A few points relating to crystallographic analyses are, however, appropriate. Icosahedral viruses are cubic solids, the highest of the Platonic solids, and they may also be described as dodecahedra. The two are complementary solids and both exhibit the same $5\bar{3}2$ symmetry. One can be inscribed within the other so that faces in one become vertices in the other, and vice versa.

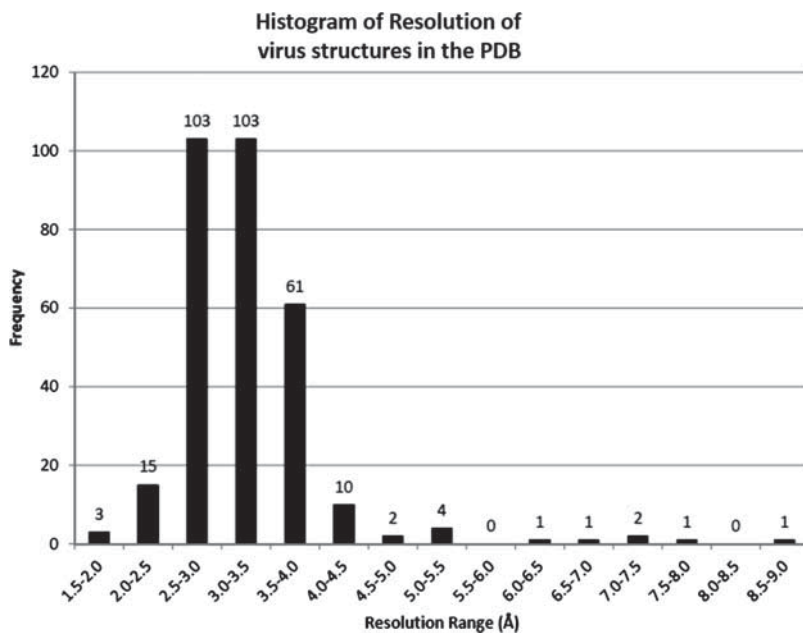


Figure 19. Histogram of the resolution limits of the X-ray crystallographic analyses of virus crystals in the PDB.

Some icosahedral viruses may in fact actually have the shape of a dodecahedron. As cubic solids they are isotropic and exhibit identical optical properties independent of direction.

Although the particles are isotropic, they can form crystals that do not exhibit isotropic properties, that is, monoclinic, orthorhombic, rhombohedral, etc., and those crystals, having different refractive indexes for different crystallographic directions, can exhibit optical effects with polarized light, including birefringence and extinction.[85,108] Because the particles making up the crystals are isotropic, however, the optical effects of virus crystals are very weak. Birefringence, for example, is dependent upon the product of the difference in the refractive index in two directions with the thickness of the crystal. To obtain any strong birefringence, therefore, it is necessary to have a large, thick virus crystal, a good fraction of a millimetre in thickness. As with all other crystals, no birefringence or extinction is possible if the virus crystal itself has cubic symmetry, or if a crystal of lower symmetry (e.g. rhombohedral and tetragonal) is viewed along an optical axis (i.e. threefold and fourfold). The end result is that optical analysis of virus crystals usually yields little reward.

As noted above, when icosahedral viruses crystallize, some of their cubic symmetry elements may be incorporated into the space group symmetry of the crystal. Thus, they are often situated on crystallographic two- or threefold axes, or at special symmetry points.[109] It is, further, not uncommon for icosahedral viruses to crystallize in cubic unit cells and reside on $2\bar{3}$ symmetry points that thereby yield the smallest possible asymmetric unit size in terms of protein subunits. Asymmetric units of icosahedral virus crystals may also be the entire virus (or even multiple n particles, e.g. PMV [46]) in which cases there are T (60) or nT (60) protein subunits as the crystallographic asymmetric unit. If residing on a rotation axis or special position, then the asymmetric unit size will be $T/2$ (60) if on a twofold axis, $T/3$ (60) on a threefold axis, $T/4$ (60) if on a $2\bar{2}2$ symmetry point, $T/6$ (60) if on a $3\bar{2}$ symmetry point, and $T/12$ (60) if positioned on a $2\bar{3}$ symmetry point.

Table 2. Space group frequencies for virus structures reported in the PDB.

Space group	Number of structures
P2 ₁	17
I222	17
C2	15
P1	14
I23	13
P2 ₁ 2 ₁ 2 (P2 ₁ 22 ₁ and P22 ₁ 2 ₁)	13
P2 ₁ 2 ₁ 2 ₁	12
H3 (R3)	12
P2 ₁ 3	9
P4 ₂ 32	9
H32 (R32)	8
P4 ₃ 2 ₁ 2	3
P6 ₃ 22	3
C222 ₁	3
P6 ₃	2
P4 ₂ 2 ₁ 2	2
P4 ₃ 22	2
P6 ₄ 22	1
P3 ₂ 2 ₁	1
F23	1
F432	1
F4 ₁ 32	1
I2 ₁ 2 ₁ 2 ₁	1
I2 ₁ 3	1
P4 ₁ 2 ₁ 2	1
P3 ₂	1

Because of the inherent symmetry of the particles, crystallographic space groups of relatively high symmetry are more common for virus crystals than for proteins and nucleic acids, but low symmetries, including P1, also frequently occur (<http://viperdb.scripps.edu/> [109]). High symmetry is to be preferred because it substantially simplifies data collection and data management, and it fixes orientation so that the entire analysis proceeds with less ambiguity (see below). Lower symmetry means that there is greater opportunity for particle averaging in the phasing and analysis stages of a structure determination.

There are 230 crystallographic space groups in total, but only 65 are possible for strictly chiral asymmetric units as is the case with biological macromolecules. Many of the possible 65, however, have not been observed for icosahedral viruses. As Table 2 shows, for 163 unique virus crystals in the PDB that served for structure determination, only 26 space groups are represented, and 9 of these only a single time. Fifteen of the 26 space groups were observed 3 or less times. The most frequently observed space groups were P2₁ and I222, which account for 21% of all space groups. The next 6 most frequent space groups C2, P1, I23, P2₁2₁2, H3, and P2₁2₁2₁ (by a large margin the most generally observed space group for globular proteins) are included, the top 8 symmetries account for 113/163 or 69% of all space groups for virus crystals. As noted already, while cubic space groups are relatively rare for biological macromolecule crystals, for virus crystals they account for 35 observations or 21% of the total.

If a particle is centred on a **32** or **23** symmetry point, then the directions of all icosahedral axes are specified. If it lies on a twofold or threefold crystallographic axis, however, then the directions of remaining particle axes must be determined to fix the orientation of the virion in the

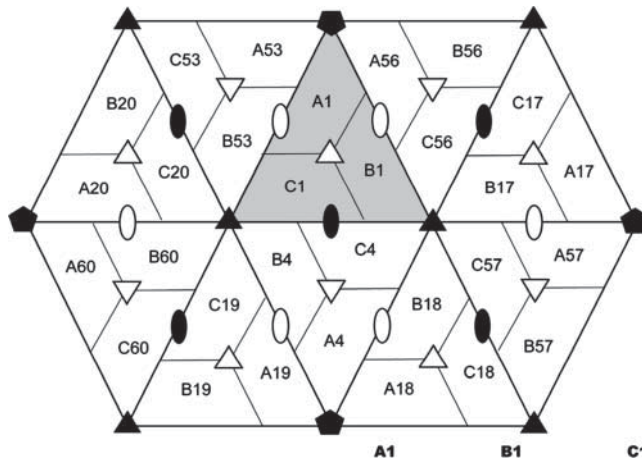


Figure 20. If a portion of the icosahedral array of a $T = 3$ virus is projected on to a plane then the interactions between adjacent A, B, and C conformers of the capsid subunits can be visualized, as it is here for the protein lattice of PMV. The numbers attached to the letters refer to the icosahedral symmetry operators that generate the subunit from the reference subunits A1, B1, and C1.

unit cell. If the virus is centred on a 222 symmetry point, then this could correspond to either of two possible orientations for the particle and, again, that ambiguity must be resolved (see below).

Because the fivefold axes of an icosahedron can never be crystallographic symmetry elements, there must always be at least $5T$ protein subunits in the asymmetric unit. This further implies that for any icosahedral virus crystal there will always be the opportunity for at least fivefold averaging within the crystallographic asymmetric unit to be exploited in phasing. For any $T > 1$ virus, however, the subunits fall into T conformational classes [2] and the T subunits do not have strictly identical conformations, though their amino acid sequences are generally the same, nor do they have identical environments. The T subunits in the icosahedral asymmetric unit are then described as being quasi-equivalent. For example, for a $T = 3$ particle, there are $T(60) = 180$ subunits, but these contain equal amounts of three quasi-equivalent variants generally denoted subunits A, B, and C (Figure 20). The three protein subunits must be treated in the analysis as different proteins.

The smallest icosahedral viruses have, of course, the lowest triangulation numbers T , such as $T = 1$, $T = 3$, $T = 4$, and $T = 7$. Beyond $T = 7$ the virions are generally too large to be addressed by single-crystal X-ray diffraction, though not, evidently by cryo-electron microscopy (article in preparation for *Crystallography Reviews* by Veesler and Johnson, and also the review by Baker et al. [106]). For $T > 1$, the principle of quasi-equivalence in which the icosahedral asymmetric unit is composed of multiple, identical proteins comes into play. Virus capsids can also have asymmetric units composed of multiple, non-identical polypeptides, or protein subunits, which results in what are termed ‘pseudo-quasi-equivalences’, in which case the symbol T is usually replaced or preceded by a p , for example $pT3$ or $p3$ [1,4,5,110,111] (<http://viperdb.scripps.edu>). Poliovirus [112] and rhinovirus [113] are prominent examples of viruses with ‘pseudo-T3 quasi-equivalence’ symmetry because of their multiple and distinct capsid polypeptide chains. The presence of ‘pseudo-quasi-equivalence’, though suggesting ominous complications, can also be precisely characterized and does not, in fact, make structure solution significantly more difficult. It simply enlarges the size of the icosahedral asymmetric unit according to the multiple polypeptide chains.

Preliminary X-ray diffraction analyses need not be carried out using a synchrotron X-ray source or with the most advanced detectors because fairly low-resolution reflections are usually

sufficient to allow determination of the unit cell parameters and the space group symmetry. Screw axis ambiguities and specification of the number of virus particles per unit cell (Z) are usually absent or are straightforward to resolve. This may not, on the other hand, be true for conventional protein crystals. Monoclinic and rhombohedral unit cells are notorious among crystallographers for their tendencies to twin.[114,115] Thus, the investigator is advised to keep a wary eye on that possibility.

To determine the true diffraction limit of a crystal of any macromolecule, including virus crystals, the crystals must be examined by X-rays at room temperature as well as cryogenic temperature, as cryo-cooling may reduce the diffraction resolution and increase mosaicity. Flash-cooling of any macromolecular crystal, and particularly large crystals (see above), inevitably produces damage, cracking or disorder that reduces the resolution of the diffraction pattern. Similarly, freezing also increases mosaicity, increases background intensity, and generally degrades the overall quality of diffraction. Crystal mounting, best done in quartz capillaries (see above), and preliminary room temperature analysis is, therefore, essential. Crystal decay from X-ray exposure can also be evaluated at room temperature, and if it is severe, then efforts to reduce it should focus on identifying cryo-crystallography conditions. There is always a trade-off between radiation damage and freezing damage, and as early on as possible that conflict needs to be resolved.

8. X-ray data collection

Recording X-ray intensities is the last truly experimental step in any crystallographic structure determination, as every procedure after that basically involves some manipulation of the X-ray amplitudes or model parameters in a computer. Thus, data collection deserves particular attention and care.[116–119] If the data are poor, subsequent steps of the analysis will be more difficult, and those steps are challenging enough with high-quality data. Recording the X-ray intensities from virus crystals, in the view of the authors, is still the most demanding part of the structure determination. Because the unit cell dimensions are several hundred or more angstroms in length and the unit cell volumes large, there are usually hundreds of thousands of independent reflections to be measured and, optimally, there will be many equivalent reflections recorded. Low crystallographic symmetry or otherwise unusually large asymmetric unit sizes intensify the challenges.

Because the total scattering of the unit cell is spread over so many intensities, any individual reflection tends to be weak and therefore associated with greater error or imprecision. Large unit cell dimensions mean very small reciprocal lattice spacings, hence the reflections are very close together and frequently difficult to resolve. The resolution limits of $T = 1$ virus crystals are generally comparable to protein crystals and several have extended to beyond 2.0 Å. Cubic crystals of SPMV yield excellent data to at least 1.9 Å resolution,[120] orthorhombic crystals of STMV diffract to at least 1.4 Å (Figure 21),[82] and crystals of STNV VLPs also to that resolution.[29] Larger $T = 3$ virus crystals do not diffract to such limits and sometimes diffract to no more than 3.5 Å at best. The histogram in Figure 19 shows the resolution limits of the structure determinations for all virus crystals in the PDB.

As discussed above, virus crystals may prove difficult to flash-cool for data collection, and even when they can be frozen, they often suffer severe damage that degrades the overall diffraction pattern. At best, freezing increases mosaicity which makes it even more difficult to resolve reflections. If it is impossible to obtain data from frozen crystals, then it becomes necessary to do so at room temperature. This places additional demand on data collection, as many crystals, generally having random orientations, are required, only short exposure times are possible, radiation damage may be severe, and clusters of only two to six frames of data must be scaled together.

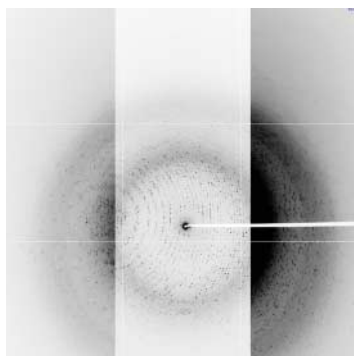


Figure 21. An oscillation image using synchrotron radiation from a frozen orthorhombic crystal of STMV is shown. Three different printing exposures from the single diffraction image are shown here so that the reflections in all resolution ranges are visible. The diffraction pattern extends to beyond 1.4 Å resolution.

As with all data collection, redundancy of measurement of reflections and their symmetry equivalents is highly desirable, but this must sometimes be reconciled with other considerations.

For structure determination of $T = 1$ particle crystals, conventional rotating anode sources have proven adequate for many cases, and the robust character of $T = 1$ viruses has allowed the use of multiple crystals and structure determination at room temperature. $T = 1$ virus crystals, including $T = 1$ particles derived from viruses, which all have diameters of 16–19 nm, such as the $T = 1$ VLP of STNV [29] or the $T = 1$ particles derived from BMV [121] and AMV,[34] have cell dimensions in the neighbourhood of 200 Å. These generally present no serious issues for data collection with rotating anode sources fitted with appropriate optical devices and using rapid detector systems such as image plates or CCD detectors.[122,123] Even multiwire detectors used in the 1990s (San Diego Multiwire Systems, San Diego, CA) proved themselves entirely adequate.

For data collection from crystals of viruses of $T > 1$, the situation is quite different (Figure 22). All of the problems alluded to above come into play. In addition, the crystals are softer, more fragile, and both mechanically and radiation sensitive. Animal and bacterial viruses such as HK 97 [124,125] epitomize the inherent problems of virus data collection. With these viruses, the use of synchrotron radiation is obligatory. Lesser sources simply do not provide sufficient intensity. The beams, furthermore, must be of very low divergence and use the best collimation available to provide the least spread of reflection intensity possible at the detector. If spot size is too great, then unacceptable numbers of reflection overlaps occur, and these are either worthless, or at best difficult to disentangle and merge. With virus crystals, beam optics [71,126] become important considerations. This is particularly so because simply reducing beam diameter also reduces X-ray flux and the volume of crystal illuminated, and therefore already weak reflections become even more so.

Spot separation is an important requirement in terms of the detector as well. In general, spot separation increases as a function of crystal to detector distance. Thus for most virus crystals, the detector is pushed back as far as possible to give the greatest crystal to detector distance. Distance, however, must be weighed against other effects. As the detector is pushed further back, the spread of reflections also increases on the detector face due to divergence, increasingly so as a function of Bragg angle, and this results in more reflection overlap. In addition, as the crystal to detector distance is increased, the angle subtended by the detector decreases and reduces the maximum Bragg angle, hence the resolution, of the recorded reflections. In practice, however, the detector is usually pushed back as far as it will go.

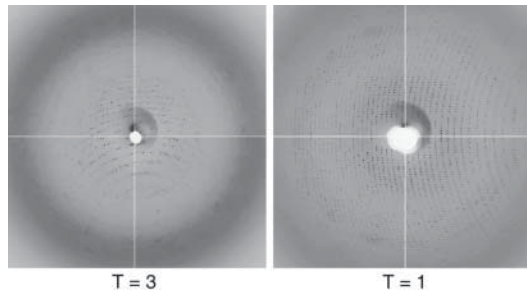


Figure 22. On the left is an oscillation image of the diffraction pattern from a frozen crystal of BMV, a 30 nm diameter, $T = 3$ virus. The diffraction limit is about 3.5 Å. On the right is an image from a frozen crystal of the 17 nm diameter, $T = 1$ particle assembled from the amino terminal abbreviated coat protein of BMV. The pattern of the $T = 1$ particle extends to about 2.7 Å. The two images are typical of those obtained from crystals of $T = 3$ and $T = 1$ particles.

With the detector at maximum distance from the crystal, it will probably be necessary to swing the detector up (or out, depending on beam line geometry) to gather reflections at high 2θ . This must be done with caution as mechanical movements in the goniostat or detector must be extremely precise in order to properly correlate the reflections ($h k l$ indexes) of high with low-resolution reflections. Reflection centres for virus crystals are separated from one another by only a few pixels on the detector. In addition, indexing of reflections at any resolution is very sensitive to the specified position of the beam centre.

Obtaining very precise coordinates for the beam centre on the detector is essential. For most protein crystals this is less important because the reciprocal lattice spacings are relatively large and the reflections on the detector well-separated. For virus crystals with large unit cell dimensions, the reflections are very close together, a few pixels of separation. Indexing of reflections is, therefore, absolutely dependent on knowing the exact beam centre. If this is in error by only two or three pixels in one or more directions, it will throw the indexing off by one or more in h , k , or l . The miseries inflicted on the individual trying to process and scale the data then become legion. More than once entire data sets have had to be entirely recollected because the beam centre was indeterminate.

The authors suggest the following expedient to locate precisely the beam centre. Before data collection on a virus crystal is initiated, or after any significant mechanical adjustment has been made, a lysozyme crystal (easily grown [58,127]) is mounted in the beam. A dozen frames or more are then quickly collected from that strongly diffracting crystal (a few minutes in total is adequate). The lysozyme data can then be quickly processed and when the unit cell is refined, the beam centre is as well. In this way, a very accurate centre point is obtained that can be trusted to support a correct indexing for the virus crystal.

Although diffraction intensities from virus crystals tend to be weak, they are not uniformly so. At low resolution, at say less than 6 Å, there may be strong intensities. In compensating for generally weak diffraction by making relatively long exposures, these strong reflections may become saturated (detector dependent) and then rejected from the data set. To obviate this, it may be necessary to recollect the low-resolution portion of the diffraction pattern with short exposure times to recapture strong reflections. This is usually done after rather than before higher resolution data collection, because low-resolution reflections are less sensitive to radiation damage. Do not ignore or dismiss the value of these strong, low-resolution reflections. Virus crystallographers will unanimously attest to their importance in the subsequent structure analysis.

Attention should be given to data collection strategy [128] to insure efficiency. Some rules are almost self-evident. If the crystal possesses a high symmetry axis (threefold, fourfold, and

sixfold), then clearly rotation around that axis provides the most rapid measurement of an entire asymmetric unit of reciprocal space. A 60° wedge about a sixfold axis is a delight, or even 90° about a fourfold axis, but neither as delightful as a 22.5° wedge in a cubic space group. Redundancy will be lacking, however, except from Friedel mates. A second orientation may be necessary about some general direction to fill in the ‘apple core’ region surrounding the first rotation axis, but only a partial data set is required for that second orientation.

If the crystal has a particularly long cell edge (correspondingly short reciprocal axis), then reflections along that direction in reciprocal space will be most difficult to resolve. If the choice presents itself, then collection by rotation about that short reciprocal axis provides the best separation. If frozen crystals cannot be used and the approach is to collect small wedges of data from many crystals and scale them together, then strategy is usually out of the question. In these circumstances, the objective must be to obtain as much data as possible from as many randomly oriented crystals as possible. That is, collect data until you run out of crystals or until the authorities show you the door.

In addition to crystal to detector distance and beam size, two other data collection parameters that deserve some thoughtful consideration are the angular increment that defines a frame of data, and the exposure time devoted to that frame. Statistical considerations of error regarding frame size suggest that small intervals or ‘fine slicing’ is preferable.[129] This is also favoured because it minimizes the overlap of reflections, usually a significant problem. On the other hand, the smaller the angular increment, the more the frames that must be collected and scaled together, and the more partial than whole reflections one will find on a frame (generally, one seldom finds an entire reflection on one frame, but the smaller the increment the more the frames over which a spot will be spread). The authors commonly use a rotation or oscillation angular increment of 0.5° unless special circumstances prevail, such as extreme overlap problems, in which case that increment may be 0.33° . With the new pixel detectors, angular increments are no longer relevant because there is continuous rotation, the shutter never closes, and there is near-continuous readout so that angular increments of as small as 0.2° are practical.[129]

Exposure time per frame also presents trade-offs. In the end, it will be determined by the diffracting power of the crystal and its sensitivity to radiation damage. Obviously, to obtain data from a weakly diffracting crystal, a frame will have to be exposed longer; the longer the exposure, the greater the magnitudes of the intensities, and the less their associated error; on the other hand, the longer the exposure, the greater the radiation damage and the fewer the frames that can be collected before the crystal becomes useless. One to two minutes of exposure per frame of 0.5° , however, might be a good starting point.

As noted above, it is necessary to use large virus crystals for data collection simply to obtain adequate X-ray intensities. Frequently, though, because of its size that crystal is considerably larger than the diameter of the X-ray beam. In that case, the beam may be directed through different, non-overlapping volumes of the crystal. This allows, in some cases, multiple clusters of frames to be collected from a single crystal at room temperature or allows an entire data set from a single frozen crystal. Thus, when collecting data from a large crystal, do not begin by shooting through the fat middle of the crystal. Begin at one end, collect there as long as you can, then move to the middle, collect, and then finally record the data from the far end. Some investigators have even mastered the technique of spiralling down a symmetry axis of a crystal.

9. X-ray data processing

Before initiating the processing of X-ray images into Lorentz-polarization (Lp) corrected structure amplitudes, it is wise to inspect the images in the set visually. Hopefully, this will produce only boredom, but it also allows one to catch unexpected, and usually unexplained, instrument or

computer glitches that may mar individual frames or a series of frames. Examination of at least some images, taken at different angular settings, sometimes exposes anomalous reflections that reveal the existence of a spur or parasite crystal, or reflections out of place that could raise the suspicion of twinning.

Examination of the images, acquired over a wide angular range, can also reveal that a crystal is cracked or split. This may not be evident in one orientation, but be pronounced in another. If the intensity distribution for a crystal is anisotropic, this may indicate that the crystal suffers from disorder in one or more directions. Problems with the diffraction data are often clearly evident by simple visual inspection that otherwise may be submerged in the statistics at later stages. There are several data processing packages that are available to the user that are well proven for virus crystal data. Prominent among these are HKL2000,[130] MOSFLM,[131] XDS,[132] and d*Trek.[118] Each may have its own specific advantages, but all have demonstrated themselves capable of handling data from virus crystals, and have been successful in yielding structure solutions. The authors favour d*Trek for its capable treatment of overlapping reflections, but we have used others as well.

Generally, processing of X-ray images requires the determination of crystal orientation and subsequent specification of the *hkl* indexes and expected position of every reflection on every image, as well as a measure of its partiality on the image.[12,133] This is dependent on the estimated mosaicity of the crystal, among other parameters. Most programs ‘track’ the images and can correct for small amounts of crystal ‘slippage’ during data collection so that the indexing and processing proceed properly. The ‘slippage’ may arise from actual crystal movement in the capillary or in the cryo-stream, stress in the fibre loop, accumulated ice or other experimental factors.

In the authors’ view, from a perspective formed from the trials and tribulations of data collection over the past 40 years, modern index assignment is little short of remarkable. No attempt will be made here to explain the algorithms and computing technology that underlie indexing, as that is done elsewhere.[133] Suffice it to say that the programs are generally capable of correctly assigning indexes and predicting precise spot positions in reciprocal space even when the reflections number in the millions (see, e.g. PMV [46]), and even when the spot separation is no more than a few pixels on the detector. When things do go wrong, however, it is most commonly at the indexing stage. As noted above, indexing is very sensitive to correct specification of the beam centre and experimental parameters. A saving grace is that most cases of mis-indexing become evident at a later stage where symmetry-related reflections are scaled and merged. Thus, a fault can be detected and corrected.

The next stage of processing after indexing and the prediction of the locations on every X-ray image of the centres of reflections is integration of the total intensity contained within the spot on the detector. This is more complicated than one might suppose. Integration depends on getting the spot centre exactly right as well as evaluating the crystal mosaicity, or reflection spread (also dependent on the divergence of the reflection), as a function of Bragg angle or position on the detector. A box or envelope of the appropriate shape and dimensions is then defined about the spot, and all of the intensities of the pixels inside the envelope summed. Following this, the background, also dependent on detector position and evaluated by separate procedures, must be subtracted from the summed intensities. Reflections are especially weak for virus crystals as discussed above, so that accurate background estimates are of crucial importance.

Integration is further complicated by the fact that with small data collection angular increments commonly used for viruses, say 0.5° , most reflections are only partially recorded on any individual image. Thus, to determine the total integrated intensity for a single observation, it is usually necessary to sum the contributions from multiple images. It is at this stage that the mosaic spread increase associated with flash-cooling of the crystals may impose itself most painfully. These difficulties are largely overcome with the application of three-dimensional profile fitting such as that implemented in the program XDS.[129,132]

Because the reciprocal lattice spacings for virus crystals are so short, and particularly with freezing, the mosaicity high, reflections tend to overlap. This is a frequent difficulty. Overlap increases with the Bragg angle so that high-resolution reflections are particularly afflicted. One approach is to simply eliminate measurements of reflections predicted to overlap. This, however, often leads to unacceptable losses of high-resolution data. Most of the programs and integration strategies, however, incorporate procedures to separate overlapping reflections and preserve I_{hkl} s.

Once indexed intensities have been obtained, the next stage in data processing is scaling and merging multiple observations of the same reflection I_{hkl} together along with symmetry equivalents. This means, at the least, I_{hkl} but generally, if the crystal symmetry is high, many other reflections as well. Anomalous pairs are seldom used in virus crystallography and Friedel mates are usually averaged. It is at this stage that outliers may be eliminated and a meaningful evaluation of the quality of the X-ray data becomes possible. There are many scaling algorithms in use and they have been treated elsewhere.[133–135] Scaling not only merges reflections, but it also smoothes out defects in the data arising from a host of sources. These include crystal shape, deterioration, beam intensity fluctuations, and absorption effects due to solvent, glass, or fibre loop. The square root of the scaled and merged intensities become the structure amplitudes, F_{hkl} s.

10. Measures of X-ray data quality

Despite the differences and cautions offered above with regard to X-ray data collection from virus crystals, assessing the quality of such data is no different than for non-virus crystals. Although there are those that claim that the relationship between model quality and data quality is uncertain,[136] intuitively, the quality of a data set should be assessed from the quality of the model derived from it, since it is doubtful that a quality model could be obtained from poor data. Although we may not be able to make an unambiguous a priori assessment of the true quality of a data set, certain statistical quantities do suggest how good it is.

The quality of a model, for example, is related to the detail seen in electron density maps, and this is dependent upon the resolution and completeness of the data.[137] Lack of completeness diminishes the effective resolution.[138] The objective should be to collect data to the diffraction limit of the crystals with as close to 100% completeness in all resolution shells as is practicable. Furthermore, the data should be as redundant as resources (i.e. beam time and suitable crystals) will allow, particularly in the highest resolution shells. High redundancy improves precision in the final averaged intensities and permits the identification of outliers through greater sampling of each reflection. It should be born in mind, however, that precision does not necessarily imply accuracy since a constant systematic error may result in high precision (low standard deviation) but inaccurate intensities.

Historically, the critical quantities that are reported with regard to data processing that have served as a summary of data quality are (1) the internal agreement (precision) of the data (expressed as R_{merge}); (2) the signal-to-noise ratio of the data (usually described as $\langle I \rangle / \langle \sigma_1 \rangle$), although some data processing programs report $\langle I \rangle / \langle \sigma_1 \rangle$); (3) the completeness, (4) the redundancy, and (5) the high-resolution limit of the data set. For each of the first four quantities, two numbers are usually given, one for the whole data set and the other for the highest resolution shell. The resolution is reported as a range for the whole data set and a range for the highest resolution shell. Referees and readers can assess the quality of the data upon which the structure is based, although model statistics (i.e. R_{work} , R_{free} , and deviations from ideal geometry) will likewise serve as an assessment of the data quality, assuming that structure solution and refinement were carried out properly. Poor model statistics but good data statistics suggest a problem with the model. Good model statistics consistent with the resolution of the data would confirm good quality set.

It has been customary to use the internal consistency and/or the signal-to-noise ratio of the highest resolution shells in specifying the resolution limit of a data set. Redundancy and completeness have likewise been used. During data processing, arbitrary targets for these four criteria may be used to set the resolution limit. Generally, the intensities or structure amplitudes are partitioned into resolution shells or bins. When a statistic calculated for the highest resolution shell does not meet the target value of one or more of these criteria, the resolution is cut to some value lower than the resolution of that bin. For example, if R_{merge} exceeds 0.5 or 0.6 or $\langle I/\sigma_I \rangle$ is less than 2.0 in the highest resolution bin (say 2.3–2.2 Å), a resolution cut-off less than the highest resolution of the bin (2.2 Å) would be applied to the data.[136,138,139] Similarly, if the average redundancy is less than 2 or the completeness is less than 50%, a resolution cut-off might also be applied. While resolution and completeness may be the primary determinants of model quality and, hence, data quality, low-resolution data sets (which are often obtained for virus data) can be evaluated by their internal consistency and signal-to-noise ratio. Even though low-resolution data sets generally result in poorer model quality, the ability to take advantage of the icosahedral symmetry of a virus through NCS map averaging tends to produce unusually good phases and, hence, maps and models that appear to be of higher resolution than otherwise would be suggested by the nominal resolution of the data.

Wlodawer et al. [138] suggested that contemporary refinement programs that employ maximum-likelihood methods, for which it is generally recommended that no data be excluded from the refinement process, allow the use of weak data without severe consequences. Hence, it would not be detrimental to process data to the maximum resolution and base the high-resolution limit in refinement on the fit of the model to the data. Furthermore, they state that ‘all reflections are very precious and should always be included, particularly at high resolution’. Under the assumption that all data are valuable, data should be processed to the highest resolution.

Ideally, data sets should have good internal agreement (precision) and high signal-to-noise ratio while maximizing the completeness, redundancy, and resolution over all resolution bins. These criteria are not independent; if high-resolution shells are discarded due to high R_{merge} or low signal-to-noise statistics, the high-resolution limit is reduced. If weak observations or outliers are rejected to improve the signal-to-noise ratio or precision, the redundancy and possibly completeness are reduced. So, data processing involves compromise among these five criteria.

The conventional measure of the internal consistency of X-ray data has been the statistic R_{merge} . Commonly used as a global indicator of data quality, R_{merge} derives from the merging of all intensity measurements of a reflection and its symmetry equivalents into a single averaged value and is given by

$$R_{\text{merge}} = \frac{\sum_h \sum_i | \langle I_h \rangle - I_{h,i} |}{\sum_h \sum_i I_{h,i}}, \quad (1)$$

where h runs through the set of unique reflections and i runs through the set of observations (including symmetry relatives) of each reflection h in the data set. Diederichs and Karplus [140] suggested that data sets with values $< 5\%$ are classified as good, 5–10% as usable, and 10–20% as marginal, and data sets with $R_{\text{merge}} > 20\%$ classified as questionable. Generally, it is expected that the lowest resolution shell have $R_{\text{merge}} < 5\%$, while the highest resolution shell should be $< 50\text{--}80\%$.[136,138,139]

As pointed out by Karplus and Diederichs,[139] the data precision statistic, R_{merge} , and the statistic for agreement of the model to the data, R_{cryst} , have very different characteristics. R_{merge} at high resolution, where reflections are weakest, tends towards infinity since the numerator is dominated by noise while the denominator tends to zero. Hence, R_{merge} should be expected to be large in the highest resolution shells. On the other hand, a value for R_{cryst} near 0.59 is representative of a random model.[141] Thus, values of R_{merge} should not be evaluated on a similar basis to R_{cryst} .

Weiss and Hilgenfeld [142] and Diederichs and Karplus [140] further pointed out that R_{merge} is inherently flawed as a global indicator of data quality because it will increase as the redundancy of the data set increases, which is somewhat counter-intuitive since more observations of an event produces a more precise description of that event. Thus, in the case of reflection intensities, the more times a reflection is measured, the more precise that measurement should be because the uncertainty or error in that measurement will decrease. Therefore, they have proposed other statistical quantities that are, so-called, redundancy independent. These include R_{meas} (also called R_{rim} for redundancy-independent merging R factor) given by

$$R_{\text{meas}} = \frac{\sum_h [N_h / (N_h - 1)]^{1/2} \sum_i | \langle I_h \rangle - I_{h,i} |}{\sum_h \sum_i I_{h,i}} \quad (2)$$

and R_{pim} (or precision-indicating merging R factor) given by

$$R_{\text{pim}} = \frac{\sum_h [1 / (N_h - 1)]^{1/2} \sum_i | \langle I_h \rangle - I_{h,i} |}{\sum_h \sum_i I_{h,i}}, \quad (3)$$

where in each case N_h is the number of observations of reflection h , and h and i are defined as for R_{merge} . R_{meas} is always larger than R_{merge} but should approach R_{merge} as the redundancy increases as shown by the tendency of the term $[N_h / (N_h - 1)]^{1/2}$ to approach 1 as N_h increases. R_{pim} can be considered an average value of the precision of redundant reflections and would have much smaller values than either R_{merge} or R_{meas} because of the $[1 / (N_h - 1)]^{1/2}$ terms.

Although the R_{merge} or R_{meas} gives some idea of the precision of a data set on a global basis, and in the individual resolution shells, and allows the data to be classified as good, usable, marginal or questionable, [140] the greater utility of these statistics may be in comparing different batches or images that produced the data set. Generally during data processing, values of R_{merge} and/or R_{meas} are calculated for each batch or each image against the whole data set. When the value for an image is considerably different than all other images, that image becomes questionable and may warrant rejection of the image. Additionally, severe decay of the diffraction due to X-ray exposure may be identified in this manner. A researcher with a data set in hand, regardless of the value of R_{merge} or R_{meas} , will attempt to solve the structure. One always does the best one can with what one has. High values for these statistics, however, may prompt the researcher to pursue better crystallization conditions or to seek better crystals with the same conditions.

In assessing the high-resolution limit by signal-to-noise ratios, the traditional target for $\langle I/\sigma_I \rangle$ in the highest resolution shell is ~ 2.0 . [138, 139] It was pointed out by Wlodawer et al., [138] however, that 48% of the structures deposited in the PDB [65] report $\langle I/\sigma_I \rangle$ of 3.0 or greater. This suggests that many structures were not determined to the maximum diffraction limit of the underlying crystals, a result of setting some arbitrary cut-off value for $\langle I/\sigma_I \rangle$ such as 3.0, or not performing a preliminary analysis of the diffraction properties of the crystals to establish the parameters for optimal data collection. An arbitrary limit for $\langle I/\sigma_I \rangle$ implies that a significant number of reflections with I/σ_I greater than the cut-off value are being discarded and, hence, potentially useful data are lost. Until the last 20–30 years, data were often eliminated from refinement by an amplitude cut-off of $F < 4.0\sigma_F$; however, with maximum-likelihood methods the recommendation is, again, to use all data. If the rule pertains to model refinement, then it should pertain to data processing as well. Once a structure is solved and refinement is initiated, an evaluation of the quality of the data in the higher resolution shells can be assessed against the model and a more practical resolution cut-off can be applied to the processed data rather than applying an arbitrary cut-off value during data processing.

Wang and Boisvert [143] demonstrated the value of weak high-resolution reflections on structure solution and refinement. They reprocessed data for a (GroEL-KMgATP)₁₄ complex that had

been truncated at 2.4 Å because the 2.4–2.3 Å resolution shell had $\langle I/\sigma_I \rangle = 1.5$. The data were reprocessed to a resolution of 2.0 Å. The number of reflections used in the subsequent refinement was about 40% greater than in the previous studies even though 143,333 reflections with $F = 0$ were excluded. The $F = 0$ reflections were, however, included in map calculations. The authors reported $\langle F/\sigma_F \rangle = 1.16$ in the 2.0–2.1 Å shell which equates to $\langle I/\sigma_I \rangle = \sim 0.58$. The final R_{work} and R_{free} values for the new refinement were 0.243 and 0.258, respectively, compared to 0.247 and 0.283 obtained using 2.4 Å resolution data with $F \geq 2.0\sigma_F$ previously reported. With the higher resolution reprocessed data, the authors were able to identify an E434A mutation, analyse probable domain motions, identify deviations from the sevenfold symmetry of the complex and nearly double the number of water molecules in the model.

A second impressive example was reported by Wang [144] involving a group II intron structure that had been truncated at 3.1 Å even though the highest resolution shell of 3.21–3.10 Å had $\langle I/\sigma_I \rangle = 3.7$. The data were reprocessed to 2.8 Å with $\langle I/\sigma_I \rangle = 0.38$ in the 2.9–2.8 Å resolution shell. The overall $\langle I/\sigma_I \rangle$ was 20.7 versus 13.9 for the previous data set. The two highest resolution shells had $R_{\text{merge}} > 100\%$. In this case, the total unique reflections increased by 37%. The final model R factors using the new data were $R_{\text{work}} = 0.196$ and $R_{\text{free}} = 0.226$ to 2.8 Å, whereas the previous model gave $R_{\text{work}} = 0.276$ and $R_{\text{free}} = 0.310$ to 3.1 Å, respectively. As a consequence, a host of additional features emerged in difference electron density maps. These two examples support the premise, suggested by others,[136,139] that measures of precision and signal-to-noise ratio are not good arbiters of the maximum resolution of a data set.

Recently, it has been suggested that the correlation coefficient of random half sets of data, designated $CC_{1/2}$, is a useful statistic for determining the high-resolution limit of a data set.[139,145] This statistic is calculated in the CCP4 programs SCALA and AIMLESS [136] and was recently added to the program XDS.[132,146] Unmerged data are randomly divided into two half sets for each unique reflection and the correlation coefficient is calculated between the average intensities of the reflections of the two sets. $CC_{1/2}$ is close to 1 at low resolution and falls sharply at resolutions near the high-resolution limit of the data as the data becomes weaker.[139] This is a more objective statistic since its calculation does not involve the uncertainties in the intensities that are more subjective since the estimation of σ_I varies among the various data processing programs.[136,139,145] Several studies [139,147,148] suggest that data sets with highest resolution shells having $CC_{1/2}$ in the range 0.1–0.2 produce better atomic models than data sets that have been truncated to a lower resolution limit. Furthermore, the work of Karplus and Diederichs [139,147] demonstrates how the $CC_{1/2}$ statistic can be a predictor of model quality through the derived statistic

$$CC^* = \left[\frac{(2CC_{1/2})}{(1 + CC_{1/2})} \right]^{1/2}, \quad (4)$$

which is an upper limit of the CC_{work} for derived models.

In summary, the data set that will yield the best model will be highly redundant, extend to a high-resolution limit characterized by a $CC_{1/2}$ statistic in the 0.1–0.2 range, and will be nearly complete in all resolution ranges with only randomly missing reflections. In other words, (1) merging R factors are of little value, especially in determining the high-resolution limit of a data set,[136,139,147,149] (2) strict signal-to-noise criteria discard useful data, degrading data quality and, consequently, model quality,[136,139,143,144,147,148] (3) highly redundant data sets are better than low redundancy sets,[147,149] (4) the $CC_{1/2}$ statistic is a better high-resolution limit indicator than previously used statistics, that is, merging R factors and $\langle I/\sigma_I \rangle$, [136,139,145,147] and (5) completeness in all resolution ranges is important, especially for structure solution, although incompleteness in the highest resolution shells only reduces the effective resolution

of the data.[138,145,147] In the words of Evans and Murshudov,[136] ‘There is no reason to suppose that cutting back the resolution of the data will improve the model.’

Occasionally, crystals may exhibit some degree of pseudosymmetry, and their X-ray diffraction patterns appear to possess higher symmetry than is actually present. Usually, however, the lower symmetry becomes evident at the scaling stage. Truly symmetry equivalent asymmetric units of reciprocal space scale together with reasonable residuals (*R* factors) comparable to those for protein crystals, at least at low and moderate resolution. Asymmetric units that only appear to be symmetry related do not; they yield markedly higher residuals. Thus, scaling statistics can be used to resolve some questions of space group.

Equivalent reflections having the same *hkl* indexes for all crystals must then be scaled and merged. The primary difference between this phase of the analysis for virus crystals in contrast to most protein crystals is the sheer number of observations and independent structure amplitudes. A second difference is that the weaker average intensity associated with virus data means that every reflection is less precisely determined and carries a greater error. As a consequence, particularly at higher resolution, statistical measures are generally inferior to those for protein crystals with smaller unit cells.

11. Determining the orientation of virus particles in the unit cell

Often a symmetry axis, or multiple symmetry axes of an icosahedral virus will be coincident with space group symmetry axes. If a single twofold axis of the particle coincides with a crystallographic twofold axis, then the asymmetric unit of the crystal will be 1/2 the particle or 30 units. For a threefold axis the asymmetric unit will be 1/3 of the particle, or 20 units. No icosahedral virus can possess a screw axis symmetry element of any order, so crystallographic screw axes can only relate entire particles in the unit cell and never result directly in an asymmetric unit of a fraction of a particle.

Icosahedral viruses may also be centred at a symmetry point, **23**, **32** or **222**, whenever they exist in the space group of a crystal. These special positions give rise to asymmetric units of 1/12, 1/6, and 1/4 (5, 10 and 15 units, respectively) of a particle. For a virus lying on a crystallographic dyad, knowing the exact orientation requires determination of its rotational angle about that axis. The same holds true if it lies on a threefold axis as well. If the unit cell has a unique origin fixed by crystallographic symmetry elements, then the position along the dyad or triad axis must also be determined.

If the virus particle is centred on a **23**, **32**, or **222** symmetry point, then its position in the cell is fixed, as is its orientation (with an ambiguity in the case of **222**). A particle, however, might lie on a twofold or threefold axis of a unit cell having a special symmetry point, but not be centred at that special position. In such a case, the position of the particle centre with respect to the special position origin must be determined. In the case of a particle centred at a **222** special position, there are two orientations of the particle that are consistent with the crystallographic symmetry. This choice must be resolved before the precise orientation of the virus can be specified.

In general, calculation of a rotation function [150,151] can resolve all questions of rotation about any axis. Because the particle has **5 3 2** symmetry, in the Eulerian rotation function, only the χ sections at 180°, 120°, and 72° need be calculated and inspected. Peaks indicating the dispositions of the 5-, 3-, and 2-fold axes are usually clear. When a rotation function is calculated, it is important that the X-ray data be complete in sampling all of reciprocal space. It is less important that weaker data are included, strong reflections alone may be adequate. Resolution too is secondary, as answers can emerge from relatively low-resolution rotation functions of 6–8 Å. It is essential that sectors of data in reciprocal space not be omitted or absent, as that may lead to erroneous rotation function results.

Once rotation angles have been determined, the only question that remains, and only for some cases, is the position of the virus centre. This also does not present a difficult problem. Packing considerations taking into account the diameter of the roughly spherical particle provide a good starting point. The particle (generally the chosen probe model, see below), in the correct orientation, can be incrementally translated in each unrestricted direction and an R factor calculated based on observed low-resolution data. By using only low, $> 10 \text{ \AA}$ structure amplitudes, a good estimate of the virus centre can be obtained. Both the coordinates of the virus centre and the angles defining the orientation of the virus, obtained from the rotation function, can then be refined precisely using higher resolution data.

It is absolutely essential that at the end of this analysis the orientation and position of the icosahedral particle be defined with precision. All subsequent operations and procedures such as isomorphous heavy atom position determination, phase extension, electron density averaging, and coordinate refinement will be completely dependent on its accuracy and precision.

12. Probe models

Crystal structures of proteins that are homologous to others of known structure are currently solved using molecular replacement,[150–152] and *de novo* structures now solved using phases based on anomalous dispersion measurements, and to a lesser extent traditional isomorphous replacement. Virus structures are not usually solved using exactly these techniques, though isomorphous replacement still has its place, and virus phasing does, at least initially (see above), use a kind of molecular substitution to obtain starting phases. With virus crystals one begins by obtaining estimates for the phases of low-resolution ($>10 \text{ \AA}$) reflections, and then, taking advantage of the symmetry within the crystallographic asymmetric unit to extend the phases to higher resolution.[153,154] The process is abetted by the high solvent volume of the crystals that allow effective solvent flattening as well.[155,156] A more detailed description of the phase determination procedure for particles having high NCS, especially viruses, is presented in an article by V. Reddy in preparation for a future issue of *Crystallography Reviews*.

To begin, however, estimates of low-resolution phases must be obtained. To accomplish this, some model, hopefully one that resembles (the closer the better) the unknown crystalline virus, is placed in the unit cell in the correct orientation and at the correct position, as determined above. Phases are then calculated from the model, and these are then used as the starting phases in a subsequent ‘boot strap’ series of procedures.

Fortunately for virus crystallographers today, a lot of advantages exist. First of all, we know that different virus species within a family closely resemble one another, particularly at low resolution. For example, TYMV and desmodium yellow mottle virus (DYMV), two tymoviruses, are almost indistinguishable,[157] as are Cowpea chlorotic mottle virus (CCMV) and BMV of the bromovirus family. Thus, if the structure of another member of the same virus family is available as a model, it is almost certain to suffice. Even if no family member is available, the amino acid sequences of the coat proteins of viruses whose structures are known can be searched for maximum amino acid identity and homology with the amino acid sequence of the crystalline virus. The VIPER data base (<http://viperd.b.scripps.edu/>) now contains well over 250 unique virus structures that have been precisely determined by X-ray diffraction. Even more models are available, though to lesser precision, based on cryo-electron microscopy, and their amino acid sequences are also known.

It is remarkable how little identity there must be between virus amino acid sequences in order for their three-dimensional structures to serve as adequate probe models for initial phase estimation. In the structure solution for PMV, for example, a particularly difficult problem because

there were two entire virus particles in the asymmetric unit (360 protein subunits), a model based on cocksfoot mottle virus (CfMV) having only about 20% amino acid identity was successful. Though it was from a different virus family, phases based on its known structure were adequate as initial phases for PMV.[46] This weak dependence on amino acid identity is undoubtedly due to the strong preservation of three-dimensional structure within the coat proteins of almost all icosahedral plant viruses. There are other classes of viruses having coat proteins of different structures including those that have large amounts of alpha helix,[25] a single jelly roll β -barrel,[158] a double jelly roll β -barrel,[159] and HK97 fold,[8,86,160] but they can usually be identified by the amino acid sequences.

Probe models based on homologous structures can also be improved before use in initial phase calculations. Differences in homologous coat proteins tend to occur as amino acid replacements, deletions or insertions, and found in extended polypeptide loops that project away from the core of secondary structure. Often these loops, being less representative of the actual virus, are simply eliminated from the model. Features that are not likely to be common between the probe model and the unknown protein, such as metal ions like Ca^{++} , should be eliminated from the probe. One can try to make appropriate amino acid substitutions from the probe model to conform to the correct amino acid sequence, but it appears unlikely that this is worth the effort. Pruning, however, may be useful. Obviously, the better the starting phases, the fewer the difficulties that will subsequently be encountered.

Although an X-ray structure-based probe model is to be preferred, such a structure may not always be available. It may be preferable then to choose a probe model of less precise structure similar to the unknown, than to use a model that is precise but does not resemble the unknown. In those cases a model based on cryo-electron microscopy may prove the best choice (see the article in preparation for a future issue of *Crystallography Reviews* by Veesler and Johnson). Cryo-electron microscopy models are generally of lower resolution ($> 8 \text{ \AA}$, though some are much better) compared with X-ray structures, but since only low-resolution phases are required, they may prove adequate, and have in a number of analyses.[107] Additional comments on the choice of probes may be found in the article by V. Reddy in preparation for a future issue of *Crystallography Reviews*.

It may be appropriate to say a few words here regarding low-resolution X-ray reflections and their value in analyses. Low-resolution reflections are generally strong, and they play the major role in defining the envelope of the virus and those spaces within the unit cell occupied by solvent. In the initial stages of phase determination that utilize chiefly low-resolution reflections, these data are of especially high significance in generating starting phases. Thus, it is wise to measure them with care and to make sure as few of them as possible are lost in data collection.

13. Heavy atoms and molecular replacement

Although extension from phases based on homologous models are generally successful in virus structure determination, it is not always the case. For a truly novel virus that may have no obvious homologues of known structure, when no adequate probe model can be found, or when, for whatever reason phase extension fails, then it is necessary to resort to traditional methods. In practice this means relying on isomorphous replacement using heavy atoms.[161,162] It is also true that isomorphous replacement phases are almost always better starting points for phase extension than those obtained from a model, and furthermore they allow that extension to be started at a higher resolution.

Isomorphous replacement was used to obtain phases for the earliest virus structures that were determined [158,159,163] and it has proven successful for many viruses that followed. In principle it operates with virus crystals in exactly the same way as with protein crystals. The target protein, the virus coat protein is no different than other proteins in composition and

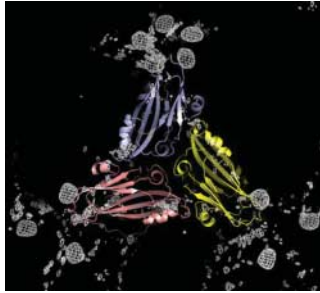


Figure 23. Shown here is a trimer of the A, B, and C protein subunits of the virus PMV with a difference Fourier map of the orthochloro mercuriphenol heavy atom derivative superimposed. The large peaks represent the binding sites of mercury atoms. Interestingly, and puzzling, even though the three different subunit conformers have the same chemical composition, the heavy atom compound binds to all 5 of the A subunits, but only the B subunits of the pseudo-hexameric B and C subunits.

it generally contains cysteine, methionine, and histidine residues that are susceptible to reaction with mercury, platinum, and silver compounds among others. They often have sites for binding lanthanides (which can replace divalent metal ions such as Mg^{++} and Ca^{++}), uranyl compounds, and other heavy atom containing organic molecules that may be attracted to low-affinity sites.[13,164]

The major difference between virus heavy atom substitution and most proteins arises from the fact that the crystallographic asymmetric units of virus crystals contain at least five and usually many more (see above) coat protein subunits. As a consequence, even when a highly selective heavy atom compound is identified that has only one or a very few reaction sites on the coat protein, there will be many substitution sites within the crystallographic asymmetric unit. The more substitution sites there are, generally the more difficult it is to determine their coordinates. Heavy atom sites are usually determined by Patterson techniques (though increasingly by direct methods), and these become increasingly complex (almost by the square) and difficult to interpret as the number of sites increases. Thus, virus crystals invariably present a complicated Patterson puzzle with heavy atoms.

The good news with virus crystals, of course, is that the multiple substitution sites within the crystallographic asymmetric unit are related by icosahedral symmetry, and that symmetry is known precisely from earlier work (see above). This makes it possible, at least in principle, to identify Patterson solutions consistent with symmetry considerations, and indeed this has been put into practice. The problem of finding multiple heavy atom substitution sites has also been reduced by new approaches to analysing Patterson maps using automated procedures,[165–167] and by the application of direct methods that have proven remarkably successful in identifying sites even when in great numbers.[168–170]

Another approach that does not require Patterson interpretation has also proven very useful. When phase extension from initial low-resolution phases based on a probe model fails for whatever reason, as it did, for example, with STMV,[55] it does not necessarily mean that the initial phases are worthless. While failing to extend to higher resolution, the phases may be adequate for calculating low-resolution $F_{HA(obs)} - F_{nat(obs)}$ difference Fourier maps on the putative heavy atom derivatives. The difference Fourier maps can then be icosahedrally averaged within the crystallographic asymmetric unit to directly reveal the heavy atom binding sites. The important point is that even marginal low-resolution phases when combined with the NCS can provide a powerful way to locate the heavy atom sites. An example where this was done, on crystalline PMV, is shown in Figure 23.

Heavy atom parameters can subsequently be refined using conventional Blow–Crick refinement,[162,171] or some corresponding algorithm, at the maximum resolution of the heavy atom derivative data. Conventional heavy atom refinement has, in recent years, largely been supplanted by maximum-likelihood approaches.[172] From the refined parameters and observed differences, phases can be determined. An advantage of these phases over those obtained from a probe model is that they suffer no model bias, that is, the resulting structure will not reflect the structure of the probe. Once heavy atom parameters have been obtained for one heavy atom derivative, then phases based on that derivative can be used alone, or in combination with probe model phases, to locate the sites for other potential derivatives. This is again done with difference Fourier syntheses.

14. A general overview of structure determination

The fundamental difference between structure determination of a conventional macromolecular crystal and a virus crystal is that, in the latter, the use of NCS, density modification, and symmetry averaging provide the essential means for phase determination and improvement.[153,154,167,173,174] To see how this is applied in practice, let us assume as a starting point that the orientation of the virus particles in the unit cell has been specified and the directions of the icosahedral axes known. Assume also that a set of initial low-resolution phases have been obtained from a probe model appropriately placed in the cell, or alternately, from isomorphous replacement. The next step is to define an envelope for the capsid that excludes the exterior solvent and may exclude the interior cavity of the virus. The envelope may be the shape of the probe model, or it may be a spherical shell the approximate thickness of the protein capsid. Selection, or definition of this is not necessarily straightforward, and all subsequent phasing, solvent flattening, and electron density averaging operations depend upon its quality. It is sometimes necessary, therefore, to try different envelopes to achieve success in obtaining accurate phases. This is apparent if one considers the pronounced variations of the exterior shapes of several different $T = 3$ viruses like those in Figure 24. A similar problem exists, but probably to a lesser extent, on the inside of a virus particle, since density modification is applied there as well. The inside surfaces, however, are more consistently smooth and uniform. An example of a capsid shell, for PMV, is shown in Figure 25.

An electron density map is calculated using the observed structure amplitudes (measured experimentally) and phases taken from the structure factors calculated from the model probe. This map will be at low resolution, perhaps at 8 Å or 10 Å. Two operations are then carried out as one. The electron density map is solvent flattened outside the envelope (both inside and outside the particle). That is, the density outside the envelope is set to some uniform low value. The solvent-flattened map is then averaged about each of the icosahedral NCS operators.[153,154,175] Presumably this map better represents reality and contains information not present in the initial map. Using the solvent flattened, averaged electron density map, new phases at the same resolution are calculated. These phases are an improvement over the earlier phases. The process is then repeated until no further changes are evident and can be monitored by calculating R factors or correlation coefficients between the observed data and the back-transformed structure amplitudes and also by the phase changes for each cycle, especially in the highest resolution bins.

At this point another set of structure factors are calculated from the solvent-flattened and symmetry-averaged electron density map, but this time the resolution of the calculated structure factors is increased by some increment Δ , usually about one reciprocal lattice point. Phases from the slightly higher resolution $F_{hkl-calc}$ are then applied to the experimentally recorded structure amplitudes and a new hybrid map is calculated at the slightly extended resolution. Again, cycles of solvent flattening, map averaging, and recalculation of maps is carried out. At the end of

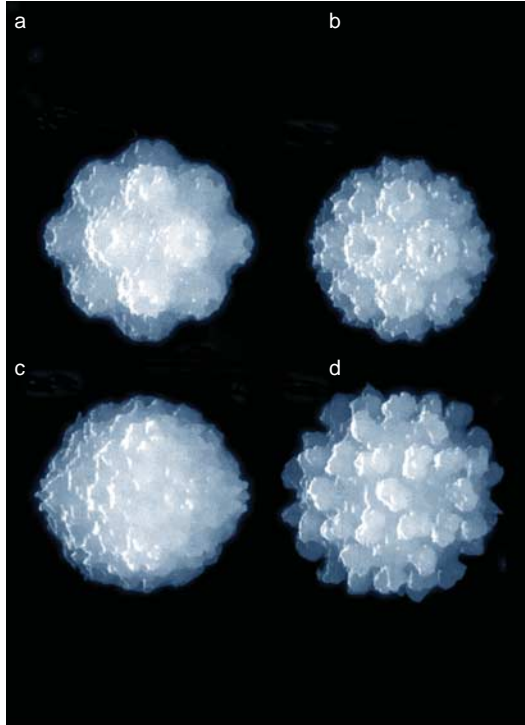


Figure 24. Shown here are surface representations of four $T = 3$ icosahedral plant viruses: (a) TYMV, (b) CCMV, (c) SBMV, and (d) TBSV. What is noteworthy is that each would require an envelope definition that is quite different from those of the others. This reveals the rather critical need to find a probe, or a known, closely related virus structure, that provides a suitable initial model and envelope for icosahedral averaging and solvent levelling during the course of phase extension from low to high resolution.

an appropriate number of cycles at the fixed resolution, the resolution is extended slightly in reciprocal space and the procedure continued. The word ‘slightly’ appears repeatedly here to caution that the extension in reciprocal space must be slow and measured. If the increment Δ is too large, then the procedure will falter. Patience here is a great virtue. It also may well be that Δ must be evaluated by trial and error to obtain the best result. Like the envelope, Δ is a critical parameter.

If fully successful then, ultimately in the resolution range of 4–3 Å, an electron density map is obtained that allows identification of polypeptide chain or some secondary structural features such as beta sheet or alpha helix. The averaging envelope should be analysed against the model to assure that it fully covers it (e.g. no loops are missing due to the envelope boundary) or that the envelope is not excessively large (tight envelopes are the most effective phase restraints under solvent flattening). At this point, structure factors and phases can be calculated from the model built into the electron density map and combined with experimental phases to produce further improved maps. One also has the option of improving the model by more conventional approaches, such as $2F_{\text{obs}} - F_{\text{calc}}$ Fourier maps, or even continuing phase extension with model-enhanced phases at each increase in resolution.

Artificial particles, or VLPs, can prove useful in improving the precision of a virus model in some cases, as can an independent, parallel structure determination of crystals of the coat protein. For example, the $T = 3$ virus BMV, though crystallized in numerous unit cells, never produced diffraction intensities much beyond about 3.5 Å resolution (Figure 22). Nevertheless, data at this marginal resolution allowed delineation of the capsid. Later, however, a $T = 1$ particle was made

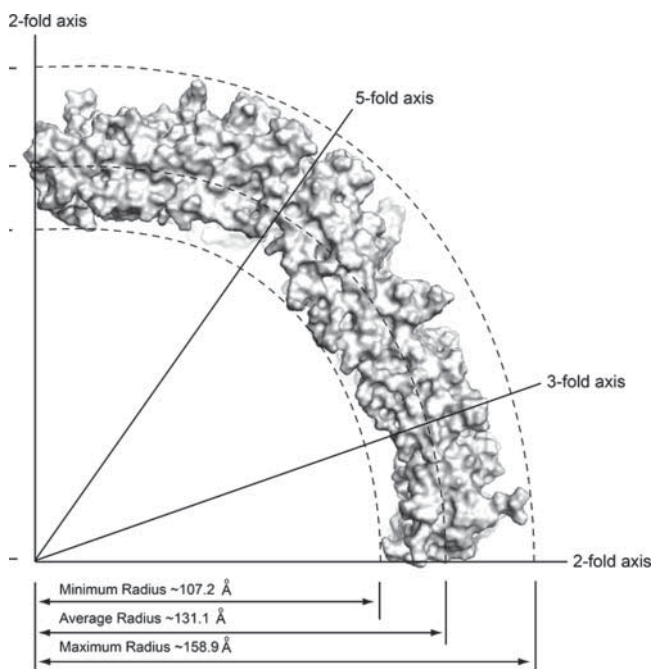


Figure 25. This diagram shows the variation in PMV capsid thickness and form and further illustrates the problem in defining a suitable envelope for phase determination. The interior space of the virion contains nucleic acid, which may be semi-ordered in some cases. This would, in principle, require different treatment in levelling than the exterior of the shell, which would be solvent.

by cleaving the amino terminal tail from the capsid protein, and crystals of these 19 nm diameter particles diffracted to about 2.9 Å. Using X-ray data from the reduced particle crystals lead to far better definition of the capsid protein, which could then be assembled according to the earlier $T = 3$, lower resolution structure.[32,52,121]

15. Refinement of virus crystal structures

The principle feature of virus crystal structure refinement that discriminates it from more conventional protein structure refinement [176–179] is the high degree of NCS from as little as 5-fold symmetry for some cubic crystals to as high as 60 fold, or even to multiples of 60 if there is more than one particle in the crystallographic asymmetric unit. Because the number of reflections available for refinement is a function of the crystallographic asymmetric unit size, but the number of independent atoms to be refined is determined by the size of the icosahedral asymmetric unit, the observation to parameter ratio is usually quite favourable. In the case of fully constrained NCS refinement, the ratio is very high and a substantially greater degree of precision is possible than for most protein crystals. For constrained refinement of STMV using 1.4 Å resolution X-ray data, for example, there were 570,721 independent reflections for 13,624 atomic parameters.[82]

Temperature factors are generally refined in viruses, as well as positional parameters, as they are in more conventional protein structures. As with the coordinates, NCS constraints or restraints must be applied to B factors as well. In general, temperature factors in virus structures have been treated isotropically, as resolution has not really permitted otherwise. B factors for virus coat proteins tend to be no higher than are observed for proteins in general, reflecting their sturdy construction and extensive intersubunit contacts. As with protein structures, the variation

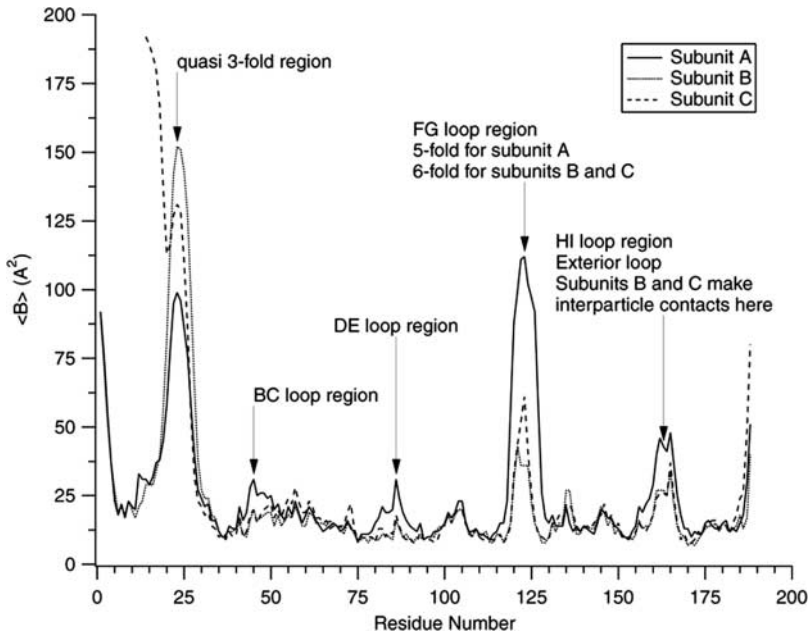


Figure 26. Shown here is a plot of refined isotropic temperature factors for the coat protein of DYMV. The high peaks represent flexible loops on the virion surface. Overall, compared to most protein structures, the temperature factors are relatively low.

in B factors is mainly indicative of the dynamics or flexibility of local regions, for example, loops and mobile termini (Figure 26). In STMV, however, 1.4 Å resolution data did permit anisotropic temperature factors to be refined.[82]

The application of fully constrained NCS refinement for virus structures assumes that the icosahedral symmetry of the virus is exact and that it is rigorously maintained when in the crystal. There is reason to doubt that this is strictly true. Crystallographic symmetry is inviolable, but the icosahedral symmetry of a virion is a consequence of biological considerations. If all protein subunits have the same amino acid sequence, and the particle is truly isotropic, as well as its chemical environment, then it should be true. On the other hand, we know that as virus particles become larger they also become softer and more deformable. Deformations from perfect sphericity might also occur when the particles are packed in a crystal lattice.

Most importantly, in a crystal lattice, local surface areas, and therefore individual protein subunits, are exposed to different environments. Some subunits may be entirely exposed to solvent and participate in no inter-particle contacts. Extended loops, or individual amino acid residues of other otherwise identical protein subunits, on the other hand may be in close contact, or form significant interactions with those on subunits of neighbouring particles in the lattice. There is no reason, therefore, to expect that NCS holds absolutely at the molecular level.

At modest resolutions of around 3 Å, or lower, it is probably safest to refine the virus structure using fully constrained refinement and exact icosahedral symmetry. This improves the observation to parameter ratio in a range where that is needed, and it represents a sound and conservative approach. At high resolution, and $T = 1$ particles have commonly been refined to beyond 2 Å Bragg spacings, most evidence suggests that restrained refinement is more appropriate. With restrained refinement exact icosahedral symmetry is not imposed, but the atoms within different subunits are allowed to deviate slightly (dependent on the nature and strengths of the restraints) from the mean position for all equivalent atoms. If very high-resolution data are available, say

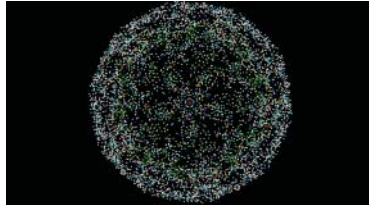


Figure 27. All water molecules that are firmly bound to either the protein or the nucleic acid of a hemisphere of STMV are shown here. The waters are colour coded according to their role in the structure [181] and make up a significant portion of the entire virus structure.

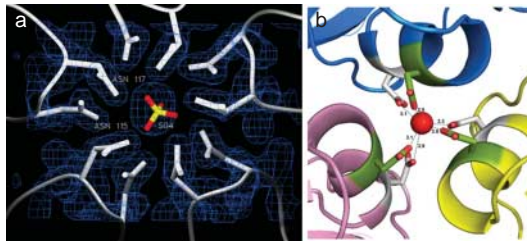


Figure 28. Both anions and cations are frequently found bound to the protein surfaces of viruses, and they utilize a variety of different ligands and modes of coordination. In (a) is a phosphate ion bound on a fivefold axis of STMV. In (b) is a calcium ion bound on a threefold axis of PMV. The ligands in (a) are the amides of 10 symmetrically disposed asparagine side chains. In (b) the calcium coordination involves six aspartic acid side groups.

beyond 1.5 Å resolution, then it may be justifiable to refine anisotropic temperature factors for some or all atoms as well.

An important feature of a virus model is the structure of the shell of ordered water molecules associated with the virus.[77,78,180] This does not include the bulk water in the crystal interstices, or at the centre of the particles, but those waters that have fixed positions by virtue of hydrogen bonding interactions with the virus. In the case of STMV, for example, it was found that ordered water molecules amounted to between 10% and 15% of the total mass of the capsid. Identifying and placing the water molecules in electron density maps is time consuming and demanding. This is particularly true with high NCS, as we find in virus crystals, where redundancies in solvent structure are difficult to detect. Nevertheless, because of their sheer number and extent, water molecules must be included. Figure 27 shows the distribution of structural water molecules in STMV. Many viruses also have ions incorporated into their structures. These may be cations, such as Mg^{++} or Ca^{++} , which are common,[6,18,19,50] or anions, as in STMV.[82] Figure 28 provides examples. These too must be identified and introduced into the model. Discriminating ions from waters, it should be noted, is frequently a challenge.

High-resolution refinement, at 1.4 Å, of at least one virus, STMV,[82] shows that as many as 30% of the amino acid side chains, principally surface residues, on the capsid proteins have alternate, or even multiple conformations. It is unlikely that amino acids on the different subunits coordinate or synchronize their conformations. This suggests that the surface features of individual viruses are constantly changing, and that no two virus particles ever appear exactly the same. It further requires that refinement, if observations allow, must take these alternate conformations into account.

Viruses, of course, contain nucleic acid, which may be single- or double-stranded RNA or DNA. The expectation is that the absence of icosahedral symmetry within the nucleic acid structure will render it invisible in electron density maps. As a consequence, although we

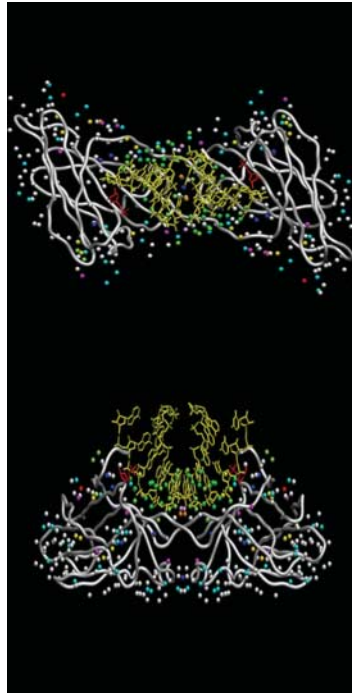


Figure 29. Orthogonal views of the double helical segment of RNA that appears at each of the 30 icosahedral dyad axes in STMV. It is bound in a cradle formed by a dimer of the coat protein. Water molecules are shown as well. At top the icosahedral twofold axis is perpendicular to the plane of the image and below it bisects the complex vertically.

acknowledge its presence, we usually cannot model it in molecular terms. It is physically 60-fold averaged when crystallization occurs, even if it is identically organized in all particles in the crystal. Currently, the only way to treat it in refinement is with something similar to a bulk solvent correction, and this is probably inadequate in most cases.

In a number of virus particles whose structures were solved by X-ray crystallography, fragments of nucleic acid, generally double helical,[182–185] but not always,[46,83,157,186] appeared in electron density maps. Figure 29 shows an example. This was because the inside of the capsid displayed sequence independent, but secondary structure-specific nucleic acid binding sites that were consistent with the icosahedral symmetry of the virion. Double helical segments of otherwise single-stranded RNA have twofold axes perpendicular to their helical axes, and in some viruses those dyads were observed to share twofold symmetry axes with the capsid and with the crystallographic symmetry elements. In some other cases (Figure 30), single strands of RNA were seen filling the cavities within pentameric or hexameric capsomeres.[182]

A measure of the degree to which segments of nucleic acid are ordered within a virion and to what extent they conform to the icosahedral symmetry of the particle is the distribution of temperature factors for the atoms in the segment. Nucleic acid structures, in general, have significantly higher temperature factors than do proteins, but nevertheless the variation among the nucleotides can be telling. In STMV, for example, helical segments bound to the interior of the protein exhibited a steep gradient. The nucleotides bound near the icosahedral dyad and in close contact with protein had unusually low B factors (Figure 31), but these increased as the segment became less ordered, or less consistent with the icosahedral symmetry. At the visible termini of the segment the temperature factors were over 150 \AA^2 .

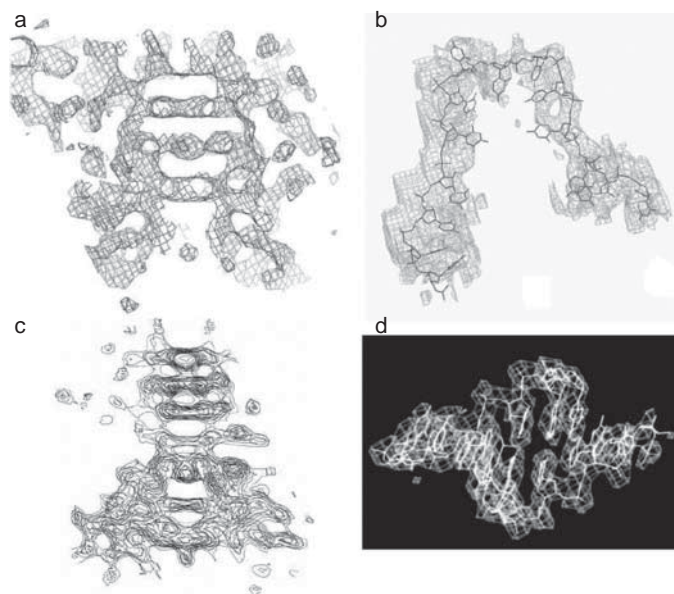


Figure 30. Drawing from [182] showing electron density for RNA in (a)–(c) TYMV and (d) STMV. Models are superimposed on the density in (b) and (d). The density in (a), (c), and (d) are helical segments. The segment in (b) is a single-stranded loop that invades the hexameric capsomeres of the tymovirus.

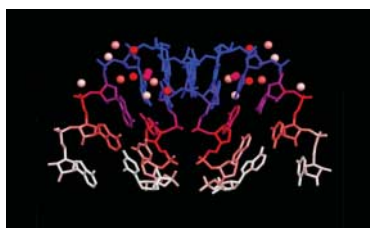


Figure 31. A segment of STMV double helical RNA of 10 nucleotides per strand where the temperature factors of the atoms are colour coded according to magnitude. Blue represents lowest values, red intermediate to high, and white very high values. It is noteworthy that the nucleic acid near the centre of the helix and in close contact with the inside of the capsid has temperature factors that are actually lower than much of the protein.

The nucleic acid must, when it is evident in the structure, be modelled and refined along with the coat protein, water and ions. This adds some complication to the refinement process, particularly when the symmetry axis of a sequence generic nucleic acid helix lies on a symmetry axis of the particle. It has also been observed that subunits having a specific conformation (e.g. the A subunits of a $T = 3$ particle) may bind nucleic acid while others do not. Protein and nucleic acids require different geometrical parameter dictionaries and the application of different restraints.

If there is one structural property of viruses that continues to remain a mystery, it is the conformation of RNA and DNA within the virion. Thus, observations of nucleic acid that provide clues are always of special interest and deserve particular attention when a virus structure is refined. In searching for nucleic acid segments, or the shadows of disordered pieces, it may be useful to take further advantage of the fact that experimentally determined phases for many viruses are so accurate that, in some cases, model phases are best traded in their favour.

An example is TYMV, a $T = 3$ plant virus. It is somewhat unique in that empty virions, devoid of RNA, are also produced naturally during infection. The full and empty virions can be separated by centrifugation and crystallized isomorphically. Crystals of each were solved independently

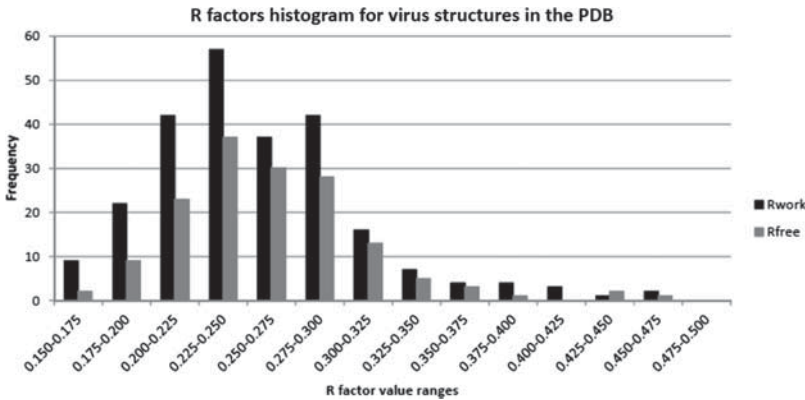


Figure 32. Shown is the histogram of the available working and free R factors for all of the X-ray crystallographic structure determinations of viruses in the PDB.

by phase extension producing corresponding sets of observed structure amplitudes plus their experimentally determined phases. It was then possible to calculate $\mathbf{F}_{\text{nat}} - \mathbf{F}_{\text{empty}}$ difference Fourier syntheses using structure factors determined exclusively experimentally. These maps were entirely free of model influence and displayed (unlike conventional difference Fourier syntheses using phases calculated from the native model) difference density belonging exclusively to ordered nucleic acid. By using such a *structure factor* difference Fourier synthesis, a substantial portion of the structured RNA in TYMV was visualized.[83]

In cases where crystallographic symmetry is low, the symmetry of the icosahedral virus often is seen as pseudosymmetry in the diffraction pattern. Even in cases where it is not obvious to the eye, statistical searches such as the rotation function can detect it. This implies that there is a correlation among intensities in the diffraction pattern due to symmetry in the structure. The eye and the rotation functions, however, examine only the structure amplitudes or intensities. If correlations can be detected among intensities, then it follows that correlations must also exist among phases as well, and of course they do. In principle, the relationship between phases of correlated structure amplitudes, as in direct methods, could provide additional phase information to a structure analysis. So far, however, this does not appear to have been explored or utilized.

The redundancy of structure that exists in real space due to icosahedral NCS has an exact analogue in reciprocal space, as is true of all properties of real and reciprocal space, though they may not always be obvious or intuitive. In the case of icosahedral viruses this means that there is a correlation between the magnitudes and phases of independent structure factors, \mathbf{F}_{hkl} , that otherwise would not exist in the absence of NCS.[174] In practice, this means that when a virus model is refined, \mathbf{F}_{hkl} s in an isolated test set participate indirectly through correlated structure amplitudes. This occurs regardless of the approach used (nonlinear least squares, maximum likelihood, etc.) when general \mathbf{F}_{hkl} s are employed to minimize the working R factor, or an equivalent residual. It follows, then, that the R_{free} is not fully independent of R and is also minimized as R is reduced.

The correlation among structure factors is seen in practice by the tendency of the R_{free} to converge to a value close to R for many virus structures that have been refined with icosahedral constraints. Hence, one must be circumspect in evaluating the quality of refinements carried out on icosahedral virus crystals and not place more weight on the value of R_{free} than it truly merits. The histogram in Figure 32 shows the distribution of R and R_{free} for the virus structures in the PDB. The average difference between R and R_{free} for all of the virus structures is significantly less than the average for most protein structures in the corresponding resolution ranges.

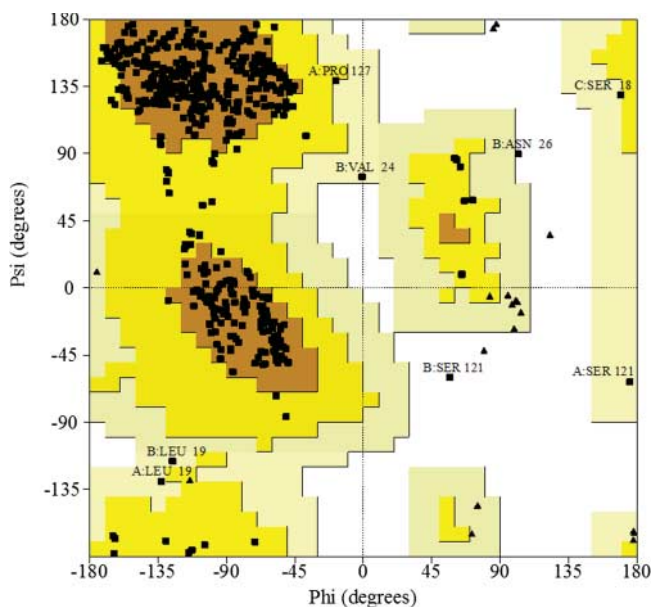


Figure 33. As with protein structures, a reliable and valuable way to assess the quality of a virus structure determination and its refinement is the Ramachandran plot that examines the distribution of phi and psi torsion angles for each amino acid. The plot shown here is for DYMV.

The problem of the correlation of R and R_{free} can be obviated to a great extent by judicious selection of the structure amplitude test set.[187] The correlation between structure factors due to NCS exists only among reflections having the same $\sin \theta$. If test sets are chosen that include all reflections within narrowly defined (thin) shells of resolution, then no correlations will be present among the reflections actually used for refinement and the test set. If this approach is applied, then the R_{free} will not track R as refinement progresses. It should be noted that in addition to statistical residuals, other measures of model quality such as Ramachandran plots (Figure 33) are equally appropriate for virus structures as for protein structures.

It has been suggested that averaging of structure using the high NCS of icosahedral viruses effectively increases the resolution of virus electron density maps. This is erroneous. Although it may be true that the use of high NCS enhances the contrast, or the level of detail, or improves the general interpretability, the true resolution of a Fourier synthesis remains limited by the maximum Bragg angle of the structure factors included in the synthesis. Resolution, of course, is strictly defined as the distance at which two scattering centres appear as two individual points. The perception that the resolution has been increased by NCS averaging in virus electron density maps arises from the pleasant fact that the phases obtained by NCS averaging and density modification for icosahedral viruses are very good in comparison with the experimental phases obtained for most protein crystals using other phase determination approaches.

Averaging using high NCS and phase extension [175] with flattening outside of a capsid envelope [156] as is currently employed for virus structure determination produces phases of inordinate accuracy, with lower average error than would probably be achieved with isomorphous replacement or even anomalous dispersion methods. It is the enhanced phase quality that produces the more detailed Fourier maps (Figure 34), but the resolution is not increased.

Criticism has, in some instances, been directed at the frequent low completion for some virus crystals of recorded data in higher resolution shells of reciprocal space. This is usually a consequence of the general problems in collecting X-ray data from virus crystals as described above.

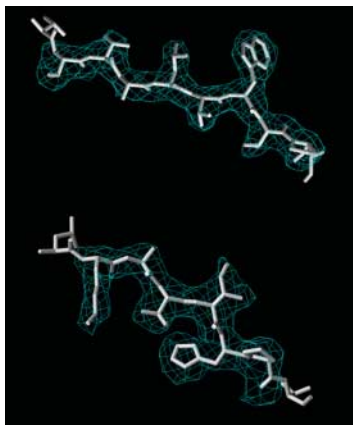


Figure 34. The quality of the phases obtained by sequential icosahedral averaging and solvent flattening are usually very good. These in turn generally yield readily interpretable electron density maps, illustrated here by the fitting of segments of polypeptide chain to short stretches of fairly typical electron density from the 2.7 Å map for DYMV.

The argument is made, for example, that 25% completion in the 3.0–2.8 Å resolution shell does not permit the claim of 2.8 Å resolution electron density maps calculated from those data. The true resolution is less. There is some truth to this, and no one would argue that a better electron density map would not be obtained if 100% of the data in the 3.0–2.8 Å resolution shell were available. It is not necessarily true, however, that the resolution is reduced significantly by only partial completion. This is again because redundancy and correlation of structure in real space is reflected as correlations between structure amplitudes and their phases in reciprocal space. Thus, if the asymmetric unit of the crystal was, for example, a quarter of a virion, implying 15-fold redundancy in real space, then 25% of the X-ray data in any resolution shell would represent an adequate sampling of reciprocal space.

Notes on contributors



Alex McPherson received a BS degree in Physics from Duke University in 1966 and a PhD in Biological Sciences from Purdue University in 1970 under the direction of Michael Rossmann. He was a Damon Runyon and then American Cancer Society postdoctoral fellow in the laboratory of Alexander Rich at MIT until 1974. He began his academic career as Assistant Professor of Biological Chemistry at the Milton S. Hershey Medical Center of the Pennsylvania State University until 1979 when he became Associate and then Professor of Biochemistry at the University of California, Riverside. He was Chairman of that department from 1986 until 1991. In 1997, he moved to the Department of Molecular Biology and Biochemistry at the University of California, Irvine where he currently resides.

Professor McPherson has written several books on macromolecular crystallization, and the analysis of macromolecular crystal structures by X-ray diffraction. He has taught in the Cold Spring Harbor course on X-ray crystallography of biological molecules for 28 years. He was also principal American investigator for macromolecular crystallization on the US Space Shuttle, Russian Space Station, and International Space Station NASA programmes. His principal interests are the study of enzyme and immunoglobulin structure, and the structures of viruses using the techniques of X-ray crystallography and atomic force microscopy.



Steven B. Larson was born in the USA (California). He received a BS degree in Chemistry and a PhD degree in Analytical Chemistry from Brigham Young University (Provo) in 1974 and 1980, respectively. He was a Robert A. Welch Postdoctoral Fellow under Stanley H. Simonsen at the University of Texas, Austin from 1980 through 1981. From 1985–1989 he headed the Analytical Instrumentation Department of the Nucleic Acid Research Institute, a subsidiary of ICN Pharmaceuticals. He converted from small molecule crystallography to macromolecular crystallography, joining the research group of Alexander McPherson in 1990 from which he retired in October 2014.

References

- [1] Chiu W, Burnett RM, Garcea R. Structural biology of viruses. Oxford: Oxford University Press; 1997.
- [2] Caspar DL, Klug A. Physical principles in the construction of regular viruses. *Cold Spring Harb Symp Quant Biol.* 1962;27:1–24.
- [3] Caspar DLD. Viral and Rickettsial infections of man. New York: Lippencott; 1965.
- [4] Casjens S. Nucleic acid packaging by viruses. In: Casjens S, editor. Virus structure and assembly. Boston, MA: Jones and Bartlett, Inc.; 1985. p. 88–95.
- [5] Rossmann MG, Johnson JE. Icosahedral RNA virus structure. *Annu Rev Biochem.* 1989;58:533–573.
- [6] Harrison SC. Principles of virus structure. In: Knipe DM, Howley PM, editors. Field's virology. Philadelphia, PA: Lippincott-Raven; 2001. p. 53–85.
- [7] Bennett A, McKenna R, Agbandje-McKenna M. A comparative analysis of the structural architecture of ssDNA viruses. *Comput Math Methods Med.* 2008;9:183–196.
- [8] Helgstrand C, Wikoff WR, Duda RL, Hendrix RW, Johnson JE, Liljas L. The refined structure of a protein catenane: the HK97 bacteriophage capsid at 3.44 Å resolution. *J Mol Biol.* 2003;334:885–899.
- [9] Abrescia NG, Cockburn JJ, Grimes JM, Sutton GC, Diprose JM, Butcher SJ, Fuller SD, San Martin C, Burnett RM, Stuart DI, Bamford DH, Bamford JK. Insights into assembly from structural analysis of bacteriophage PRD1. *Nature.* 2004;432:68–74.
- [10] Arnold E, Himmel DM, Rossmann MG. International tables for crystallography. Vol. F, crystallography of biological macromolecules. 2nd ed. New York: John Wiley; 2012.
- [11] McPherson A. Introduction to the crystallization of biological macromolecules. In: DeLucas L, editor. Membrane protein crystallization. Vol. 63. Burlington, MA: Elsevier; 2009. p. 5–23.
- [12] Rupp B. Biomolecular crystallography: principles, practice and applications to structural biology. New York: John Wiley Co.; 2009.
- [13] Blundell TL, Johnson LN. Protein crystallography. New York: Academic Press; 1976.
- [14] Rhodes G. Crystallography made crystal clear. New York: Academic Press; 2002.
- [15] Rossmann MG. Virus crystallography and structural virology: a personal perspective. *Crystallogr Rev.* 2014. doi:10.1080/0889311X.2014.957282.
- [16] Laliberte JF, Moffett P, Sanfacon H, Wang A, Nelson RS, Schoelz JE. e-Book on plant virus infection – a cell biology perspective. *Front Plant Sci.* 2013;4:203. doi:10.3389/fpls.2013.00203.
- [17] Solovyev AG, Savenkov EI. Factors involved in the systemic transport of plant RNA viruses: the emerging role of the nucleus. *J Exp Bot.* 2014;65:1689–1697.
- [18] Kaper JM. The chemical basis of virus structure, dissociation and reassembly. In: Neuberger A, Tatum EL, editors. The frontiers of biology series. Amsterdam: North-Holland; 1975. p. 1–485.
- [19] Hull R. Matthews' plant virology. 4th ed. Amsterdam: Elsevier Ltd; 2002.
- [20] Knipe DM, Howley PM. Field's virology. 5th ed. Amsterdam: Lippencott-Wilkins; 2007.
- [21] Munshi S, Liljas L, Johnson JE. Structure determination of Nudaurelia capensis omega virus. *Acta Cryst.* 1998;D54:1295–1305.
- [22] Wery JP, Reddy VS, Hosur MV, Johnson JE. The refined three-dimensional structure of an insect virus at 2.8 Å resolution. *J Mol Biol.* 1994;235:565–586.
- [23] Tate J, Liljas L, Scotti P, Christian P, Lin T, Johnson JE. The crystal structure of cricket paralysis virus: the first view of a new virus family. *Nat Struct Biol.* 1999;6:765–774.
- [24] Federici BA. Isolation of an iridovirus from two terrestrial isopods, the pill bug, *Armadillidium vulgare*, and the sow bug, *Porcellio dilatatus*. *J Invertebr Pathol.* 1980;36:373–381.
- [25] Grimes JM, Burroughs JN, Gouet P, Diprose JM, Malby R, Zientara S, Mertens PP, Stuart DI. The atomic structure of the bluetongue virus core. *Nature.* 1998;395:470–478.

- [26] Hu L, Chow DC, Patton JT, Palzkill T, Estes MK, Prasad BV. Crystallographic analysis of rotavirus NSP2-RNA complex reveals specific recognition of 5' GG sequence for RTPase activity. *J Virol.* 2012;86:10547–10557.
- [27] Stewart PL, Burnett RM. Adenovirus structure by X-ray crystallography and electron microscopy. *Curr Top Microbiol Immunol.* 1995;199:25–38.
- [28] Birtley JR, Curry S. Crystallization of foot-and-mouth disease virus 3C protease: surface mutagenesis and a novel crystal-optimization strategy. *Acta Cryst.* 2005;D61:646–650.
- [29] Lane SW, Dennis CA, Lane CL, Trinh CH, Rizkallah PJ, Stockley PG, Phillips SE. Construction and crystal structure of recombinant STNV capsids. *J Mol Biol.* 2011;413:41–50.
- [30] Yang L, Song Y, Li X, Huang X, Liu J, Ding H, Zhu P, Zhou P. HIV-1 virus-like particles produced by stably transfected *Drosophila* S2 cells: a desirable vaccine component. *J Virol.* 2012;86:7662–7676.
- [31] Kuznetsov YG, Ulbrich P, Haubova S, Ruml T, McPherson A. Atomic force microscopy investigation of Mason-Pfizer monkey virus and human immunodeficiency virus type 1 reassembled particles. *Virology.* 2007;360:434–446.
- [32] Lucas RW, Kuznetsov YG, Larson SB, McPherson A. Crystallization of brome mosaic virus (BMV) and $T = 1$ brome mosaic virus particles following a structural transition. *Virology.* 2001;286:290–303.
- [33] Cuillel M, Jacrot B. A $T = 1$ capsid formed by the protein of brome mosaic virus in the presence of trypsin. *Virology.* 1981;110:63–72.
- [34] Kumar A, Reddy VS, Yusibov V, Chipman PR, Hata Y, Fita I, Fukuyama K, Rossmann MG, Loesch-Fries LS, Baker TS, Johnson JE. The structure of alfalfa mosaic virus capsid protein assembled as a $T = 1$ icosahedral particle at 4.0-Å resolution. *J Virol.* 1997;71:7911–7916.
- [35] Sainsbury F, Saxena P, Aljabali AA, Saunders K, Evans DJ, Lomonosoff GP. Genetic engineering and characterization of Cowpea mosaic virus empty virus-like particles. *Methods Mol Biol.* 2014;1108:139–153.
- [36] Kuznetsov YG, Zhang M, Menees TM, McPherson A, Sandmeyer S. Investigation by atomic force microscopy of the structure of Ty3 retrotransposon particles. *J Virol.* 2005;79:8032–8045.
- [37] Larsen LSZ, Kuznetsov YG, McPherson A, Hatfield GW, Sandmeyer S. TY3 GAG3 protein forms ordered particles in *Escherichia coli*. *Virology.* 2008;370:223–227.
- [38] Martin SJ. The biochemistry of viruses. Cambridge, UK: Cambridge University Press; 1978.
- [39] Green AA, Hughes WL. Protein fractionation on the basis of solubility in aqueous solutions of salts and organic solvents. *Methods Enzymol.* 1955;1:67–90.
- [40] Hofmeister F. Zur Lehre von der Wirkung der Salze Nauyen-Schmiedebergs. *Arch Exp Pathol Pharmacol.* 1888;24:247–260.
- [41] Kay LE. W. M. Stanley's crystallization of the tobacco mosaic virus, 1930–1940. *Isis.* 1986;77:450–472.
- [42] Matthews RC. Plant virology. 4th ed. New York: Academic Press; 2001.
- [43] Jaspars EMJ. Interaction of alfalfa mosaic virus nucleic acid and protein. In: Davis JW, editor. *Molecular plant virology*. Boca Raton, FL: CRC Press, Inc; 1985. p. 155–221.
- [44] McPherson A. Crystallization of proteins from polyethylene glycol. *J Biol Chem.* 1976;251:6300–6303.
- [45] Makino DL, Day J, Larson SB, McPherson A. Investigation of RNA structure in satellite panicum mosaic virus. *Virology.* 2006;351:420–431.
- [46] Makino DL, Larson SB, McPherson A. The crystallographic structure of panicum mosaic virus (PMV). *J Struct Biol.* 2013;181:37–52.
- [47] Kaper JM. Molecular organization and stabilizing forces of simple RNA viruses. V. The role of lysyl residues in the stabilization of cucumber mosaic virus strain S. *Virology.* 1976;71:185–198.
- [48] Speir JA, Bothner B, Qu C, Willits DA, Young MJ, Johnson JE. Enhanced local symmetry interactions globally stabilize a mutant virus capsid that maintains infectivity and capsid dynamics. *J Virol.* 2006;80:3582–3591.
- [49] Kuznetsov YG, Larson SB, Day J, Greenwood A, McPherson A. Structural transitions of satellite tobacco mosaic virus particles. *Virology.* 2001;284:223–234.
- [50] Aramayo R, Merigoux C, Larquet E, Bron P, Perez J, Dumas C, Vachette P, Boisset N. Divalent ion-dependent swelling of tomato bushy stunt virus: a multi-approach study. *Biochim Biophys Acta.* 2005;1724:345–354.
- [51] Wang L, Lane LC, Smith DL. Detecting structural changes in viral capsids by hydrogen exchange and mass spectrometry. *Protein Sci.* 2001;10:1234–1243.
- [52] Lucas RW, Larson SB, McPherson A. The crystallographic structure of brome mosaic virus. *J Mol Biol.* 2002;317:95–108.

- [53] Makino DL, Larson SB, McPherson A. Preliminary analysis of crystals of panicum mosaic virus (PMV) by X-ray diffraction and atomic force microscopy. *Acta Cryst.* 2005;D61:173–179.
- [54] Argos P, Johnson JE. Chemical stability in simple spherical plant viruses. In: Jurnak FA, McPherson A, editors. *Biological macromolecules and assemblies*. Vol. 1, virus structures. New York: John Wiley and Sons; 1984. p. 1–32.
- [55] Larson SB, Koszelak S, Day J, Greenwood A, Dodds JA, McPherson A. Three-dimensional structure of satellite tobacco mosaic virus at 2.9 Å resolution. *J Mol Biol.* 1993;231:375–391.
- [56] Jaspars EM. Plant viruses with a multipartite genome. *Adv Virus Res.* 1974;19:37–149.
- [57] Van Vloten-Doting L. Advantages of multipartite genomes of ss-RNA plant viruses in nature, for research, and for genetic engineering. *Plant Mol Biol Rep.* 1983;1:55–60.
- [58] McPherson A. *Crystallization of biological macromolecules*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 1999.
- [59] McPherson A. Macromolecular crystallization in the structural genomics era. *J Struct Biol.* 2003;142:1–2.
- [60] McPherson A. Introduction to protein crystallization. *Methods Companion Methods Enzymol.* 2004;34:254–265.
- [61] McPherson A, Gavira JA. Introduction to protein crystallization. *Acta Cryst.* 2014;F70:2–20.
- [62] McPherson A. *The preparation and analysis of protein crystals*. New York: John Wiley and Sons; 1982.
- [63] Bergfors TM, editor. *Protein crystallization: techniques, strategies and tips*. La Jolla, CA: International University Line; 1999.
- [64] Ducruix A, Giegé R. *Crystallization of nucleic acids and proteins*. 2nd ed. Oxford: Oxford University Press; 2000.
- [65] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE. The Protein Data Bank. *Nucleic Acids Res.* 2000;28:235–242.
- [66] McPherson A, Cudney B. Searching for silver bullets: an alternative strategy for crystallizing macromolecules. *J Struct Biol.* 2006;156:387–406.
- [67] McPherson A, Cudney B. Optimization of crystallization conditions for biological macromolecules. *Acta Cryst F*. Forthcoming.
- [68] Garman EF, Schneider TR. Macromolecular cryocrystallography. *J Appl Cryst.* 1997;30:211–237.
- [69] Hope H, Parkin S. Cryocrystallography: introduction to cryocrystallography. In: Arnold E, Himmel DM, Rossmann MG, editors. *International tables for crystallography*. Vol. F, crystallography of biological macromolecules. 2nd ed. New York: Wiley Co.; 2012. p. 241–248.
- [70] Garman EF. Cryocrystallography: radiation damage. In: Arnold E, Himmel DM, Rossmann MG, editors. *International tables for crystallography*. Vol. F, crystallography of biological macromolecules. 2nd ed. New York: Wiley; 2012. p. 256–261.
- [71] Helliwell JR. *Macromolecular crystallography with synchrotron radiation*. Cambridge, UK: Cambridge University Press; 1992.
- [72] Malkin AJ, Kuznetsov YG, McPherson A. Defect structure of macromolecular crystals. *J Struct Biol.* 1996;117:124–137.
- [73] McPherson A, Kuznetsov YG. Mechanisms, kinetics, impurities and defects: consequences in macromolecular crystallization. *Acta Cryst.* 2014;F70:384–403.
- [74] Malkin AJ, Plomp M, McPherson A. Application of atomic force microscopy to studies of surface processes in virus crystallization and structural biology. *Acta Cryst.* 2002;D58:1617–1621.
- [75] Malkin AJ, Plomp M, McPherson A. Unraveling the architecture of viruses by high-resolution atomic force microscopy. In: Lieberman PM, editor. *Virus structure and imaging, DNA viruses, methods and protocols*. Totowa, NJ: Humana Press; 2004. p. 85–108.
- [76] Malkin AJ, Kuznetsov YG, Land TA, DeYoreo JJ, McPherson A. Mechanisms of growth for protein and virus crystals. *Nat Struct Biol.* 1995;2:956–959.
- [77] Frey M. Water structure associated with proteins and its role in crystallization. *Acta Cryst.* 1994;D50:663–666.
- [78] Mattos C, Ringe D. Structural analysis and classification: solvent structure. In: Arnold E, Himmel DM, Rossmann MG, editors. *International tables for crystallography*. Vol. F, crystallography of biological macromolecules. 2nd ed. New York: Wiley; 2012. p. 800–820.
- [79] Gilliland GL, Tung M, Ladner J. The biological macromolecule crystallization database and NASA protein crystal growth archive. *J Res Natl Inst Stand Technol.* 1996;101:309–320.
- [80] Rosenberger A. *Fundamentals of crystal growth*. Berlin: Springer-Verlag; 1979.
- [81] Chernov AA, editor. *Modern crystallography, III, crystal growth*. Berlin: Springer-Verlag; 1984.

- [82] Larson SB, Day JS, McPherson A. Satellite tobacco mosaic virus refined to 1.4 Å resolution. *Acta Cryst.* 2014;D70:2316–2330.
- [83] Larson SB, Lucas RW, Greenwood A, McPherson A. The RNA of turnip yellow mosaic virus exhibits icosahedral order. *Virology.* 2005;334:245–254.
- [84] King MV. An efficient method for mounting wet protein crystals for X-ray studies. *Acta Cryst.* 1954;7:601–602.
- [85] Carrell HL, Glusker JP. Crystal properties and handling: crystal morphology, optical properties of crystals and crystal mounting. In: Arnold E, Himmel DM, Rossmann MG, editors. *International tables for crystallography. Vol. F, crystallography of biological macromolecules.* 2nd ed. New York: Wiley; 2012. p. 145–151.
- [86] Wikoff WR, Schildkamp W, Johnson JE. Increased resolution data from a large unit cell crystal collected at a third-generation synchrotron X-ray source. *Acta Cryst.* 2000;D56:890–893.
- [87] Banatao DR, Cascio D, Crowley CS, Fleissner MR, Tienson HL, Yeates TO. An approach to crystallizing proteins by synthetic symmetrization. *Proc Natl Acad Sci USA.* 2006;103:16230–16235.
- [88] Laganowsky A, Zhao M, Soriaga AB, Sawaya MR, Cascio D, Yeates TO. An approach to crystallizing proteins by metal-mediated synthetic symmetrization. *Protein Sci.* 2011;20:1876–1890.
- [89] Padilla JE, Colovos C, Yeates TO. Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. *Proc Natl Acad Sci USA.* 2001;98:2217–2221.
- [90] Bamford DH, Burnett RM, Stuart DI. Evolution of viral structure. *Theor Popul Biol.* 2002;61:461–470.
- [91] Malkin AJ, Cheung J, McPherson A. Crystallization of satellite tobacco mosaic virus I. Nucleation phenomena. *J Cryst Growth.* 1993;126:544–554.
- [92] Malkin A, McPherson A. Light scattering investigations of protein and virus crystal growth: ferritin, apoferritin and satellite tobacco mosaic virus. *J Cryst Growth.* 1993;128:1232–1235.
- [93] Kuznetsov YG, Malkin AJ, Lucas RW, McPherson A. Atomic force microscopy studies of icosahedral virus crystal growth. *Colloids Surf B Biointerfaces.* 2000;19:333–346.
- [94] Kuznetsov YG, Konnert J, Malkin AJ, McPherson A. The advancement and structure of growth steps on thaumatin crystals visualized by atomic force microscopy at molecular resolution. *Surf Sci.* 1999;440:69–80.
- [95] McPherson A, Malkin AJ, Kuznetsov Yu G. Atomic force microscopy in the study of macromolecular crystal growth. *Annu Rev Biophys Biomol Struct.* 2000;29:361–410.
- [96] Buckley HE. *Crystal growth.* London: John Wiley and Sons; 1951.
- [97] Burton WK, Cabrera N, Frank FC. The growth of crystals and the equilibrium structure of their surfaces. *Philos Trans R Soc.* 1951;A243:299–358.
- [98] McPherson A, Malkin AJ, Kuznetsov YG. The science of macromolecular crystallization. *Structure.* 1995;3:759–768.
- [99] Vekilov PG, Lin H, Rosenberger F. Unsteady crystal growth due to step-bunch cascading. *Phys Rev Lett.* 1997;E55:3202–3209.
- [100] Kuznetsov YG, Malkin AJ, McPherson A. Self-repair of biological fibers catalyzed by the surface of a virus crystal. *Proteins.* 2001;44:392–396.
- [101] McPherson A, Malkin A, Kuznetsov YG, Koszelak S. Incorporation of impurities into macromolecular crystals. *J Cryst Growth.* 1996;168:74–92.
- [102] Malkin AJ, Kuznetsov YG, McPherson A. Incorporation of microcrystals by growing protein and virus crystals. *Proteins.* 1996;24:247–252.
- [103] Malkin AJ, Kuznetsov YG, McPherson A. An in situ investigation of catalase crystallization. *Surf Sci.* 1997;393:95–107.
- [104] Kuznetsov YG, Makino DL, Malkin AJ, McPherson A. The incorporation of large impurity into virus crystals. *Acta Cryst.* 2005;D61:720–723.
- [105] Cowley JM. *Diffraction physics.* 2nd ed. Amsterdam: North Holland; 1984.
- [106] Malkin AJ, McPherson A. Novel mechanisms for defect formation and surface molecular processes in virus crystallization. *J Phys Chem.* 2002;106:6718–6722.
- [107] Baker TS, Olson NH, Fuller SD. Adding the third dimension to virus life cycles: three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. *Microbiol Mol Biol Rev.* 1999;63:862–922.
- [108] Wood EA. The relation between the symmetry of a crystal and the symmetry of its physical properties. *Crystals and light: an introduction for optical crystallography.* 2nd ed. New York: Dover; 1977. p. 69–78.

- [109] Carrillo-Tripp M, Shepherd CM, Borelli IA, Venkataraman S, Lander G, Natarajan P, Johnson JE, Brooks CL, 3rd Reddy VS. VIPERdb2: an enhanced and web API enabled relational database for structural virology. *Nucleic Acids Res.* 2009;37:D436–D442.
- [110] Johnson JE, Speir JA. Quasi-equivalent viruses: a paradigm for protein assemblies. *J Mol Biol.* 1997;269:665–675.
- [111] Izaac A, Schall CA, Mueser TC. Crystallization optimum solubility screening: using crystallization results to identify the optimal buffer for protein crystal formation. *Acta Cryst.* 2005;F61:1035–1038.
- [112] Hogle JM, Chow M, Filman DJ. Three-dimensional structure of poliovirus at 2.9 Å resolution. *Science.* 1985;229:1358–1365.
- [113] Hadfield AT, Lee W, Zhao R, Oliveira MA, Minor I, Rueckert RR, Rossmann MG. The refined structure of human rhinovirus 16 at 2.15 Å resolution: implications for the viral life cycle. *Structure.* 1997;5:427–441.
- [114] Yeates TO. Detecting and overcoming crystal twinning. *Methods Enzymol.* 1997;276:344–358.
- [115] Ramagopal UA, Dauter M, Dauter Z. Phasing on anomalous signal of sulfurs: what is the limit? *Acta Cryst.* 2003;D59:1020–1027.
- [116] Dauter Z, Wilson KS. X-ray data collection: principles of monochromatic data collection. In: Arnold E, Himmel DM, Rossmann MG, editors. *International tables for crystallography. Vol. F, crystallography of biological macromolecules.* 2nd ed. New York: Wiley; 2012. p. 211–230.
- [117] Dauter Z. Data-collection strategies. *Acta Cryst.* 1999;D55:1703–1717.
- [118] Pflugrath JW. The finer things in X-ray diffraction data collection. *Acta Cryst.* 1999;D55:1718–1725.
- [119] Garman E, Sweet RM. X-ray data collection from macromolecular crystals. *Methods Mol Biol.* 2007;364:63–94.
- [120] Ban N, McPherson A. The structure of satellite panicum mosaic virus at 1.9 Å resolution. *Nat Struct Biol.* 1995;2:882–890.
- [121] Larson SB, Lucas RW, McPherson A. Crystallographic structure of the $T = 1$ particle of brome mosaic virus. *J Mol Biol.* 2005;346:815–831.
- [122] Gruner SM, Eikenberry EF, Tate MW. X-ray detectors: comparison of X-ray detectors. In: Arnold E, Himmel DM, Rossmann MG, editors. *International tables for crystallography. Vol. F, crystallography of biological macromolecules.* 2nd ed. New York: Wiley; 2012. p. 177–182.
- [123] Tate MW, Eikenberry EF, Gruner SM. X-ray detectors: CCD detectors. In: Arnold E, Himmel DM, Rossmann MG, editors. *International tables for crystallography. Vol. F, crystallography of biological macromolecules.* 2nd ed. New York: Wiley; 2012. p. 183–188.
- [124] Wikoff WR, Duda RL, Hendrix RW, Johnson JE. Crystallographic analysis of the dsDNA bacteriophage HK97 mature empty capsid. *Acta Cryst.* 1999;D55:763–771.
- [125] Gan L, Speir JA, Conway JF, Lander G, Cheng N, Firek BA, Hendrix RW, Duda RL, Liljas L, Johnson JE. Capsid conformational sampling in HK97 maturation visualized by X-ray crystallography and cryo-EM. *Structure.* 2006;14:1655–1665.
- [126] Arndt UW. Radiation sources and optics: X-ray sources. In: Arnold E, Himmel DM, Rossmann MG, editors. *International tables for crystallography. Vol. F, crystallography of biological macromolecules.* 2nd ed. New York: Wiley; 2012. p. 159–167.
- [127] Alderton G, Fevold HL. Direct crystallization of lysozyme from egg white and some crystalline salts of lysozyme. *J Biol Chem.* 1946;164:1–5.
- [128] Vickovic I, Kalk KH, Drenth J, Dijkstra BW. An optimal strategy for X-ray data collection on macromolecular crystals with position-sensitive detectors. *J Appl Cryst.* 1994;27:791–793.
- [129] Mueller M, Wang M, Schulze-Briese C. Optimal fine φ -slicing for single-photon-counting pixel detectors. *Acta Cryst.* 2012;D68:42–56.
- [130] Otwinowski Z, Minor W. Processing of X-ray diffraction data collected in oscillation mode. In: Carter JW, Sweet RM, editors. *Macromolecular crystallography.* San Diego, CA: Academic Press; 1997. p. 307–326.
- [131] Leslie AGW. The integration of macromolecular diffraction data. *Acta Cryst.* 2006;D62:48–57.
- [132] Kabsch W. XDS. *Acta Cryst.* 2010;D66:125–132.
- [133] Rossmann MG. Data processing: automatic indexing of oscillation images. In: Arnold E, Himmel DM, Rossmann MG, editors. *International tables for crystallography. Vol. F, crystallography of biological macromolecules.* 2nd ed. New York: Wiley; 2012. p. 263–265.
- [134] Kabsch W. Data processing: integration, scaling, space group assignment and post refinement. In: Arnold E, Himmel DM, Rossmann MG, editors. *International tables for crystallography. Vol. F, crystallography of biological macromolecules.* 2nd ed. New York: Wiley; 2012. p. 272–281.

- [135] Van Beek CG, Bolotovskiy R, Rossmann MG. Data processing: the use of partially recorded reflections for post refinement, scaling, and averaging X-ray diffraction data. In: Arnold E, Himmel DM, Rossmann MG, editors. *International tables for crystallography*. Vol. F, crystallography of biological macromolecules. 2nd ed. New York: Wiley; 2012. p. 296–303.
- [136] Evans PR, Murshudov GN. How good are my data and what is the resolution? *Acta Cryst.* 2013;D69:1204–1214.
- [137] Stenkamp RE, Jensen LH. Resolution revisited: limit of detail in electron density maps. *Acta Cryst.* 1984;A40:251–254.
- [138] Wlodawer A, Minor W, Dauter Z, Jaskolski M. Protein crystallography for aspiring crystallographers or how to avoid pitfalls and traps in macromolecular structure determination. *FEBS J.* 2013;280:5705–5736.
- [139] Karplus PA, Diederichs K. Linking crystallographic model and data quality. *Science.* 2012;336:1030–1033.
- [140] Diederichs K, Karplus PA. Improved R-factors for diffraction data analysis in macromolecular crystallography. *Nat Struct Biol.* 1997;4:269–275.
- [141] Wilson AJC. Largest likely values for the reliability index. *Acta Cryst.* 1950;3:397–398.
- [142] Weiss MS, Hilgenfeld R. On the use of the merging R factor as a quality indicator for X-ray data. *J Appl Cryst.* 1997;30:203–205.
- [143] Wang J, Boisvert DC. Structural basis for GroEL-assisted protein folding from the crystal structure of (GroEL-KMgATP)₁₄ at 2.0 Å resolution. *J Mol Biol.* 2003;327:843–855.
- [144] Wang J. Inclusion of weak high-resolution X-ray data for improvement of a group II intron structure. *Acta Cryst.* 2010;D66:988–1000.
- [145] Evans PR. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Cryst.* 2011;D67:282–292.
- [146] Kabsch W. Integration, scaling, space-group assignment and post-refinement. *Acta Cryst.* 2010;D66:133–144.
- [147] Diederichs K, Karplus PA. Better models by discarding data? *Acta Cryst.* 2013;D69:1215–1222.
- [148] Wang J, Wing RA. Diamonds in the rough: a strong case for the inclusion of weak-intensity X-ray diffraction data. *Acta Cryst.* 2014;D70:1491–1497.
- [149] Weiss MS. Global indicators of X-ray data quality. *J Appl Cryst.* 2001;34:130–135.
- [150] Rossmann MG, editor. *The molecular replacement method*. London: Gordon and Breach; 1972.
- [151] Rossmann MG, Blow DM. The detection of subunits within the crystallographic asymmetric unit. *Acta Cryst.* 1962;15:24–31.
- [152] Blow DM. Molecular replacement: noncrystallographic symmetry. In: Arnold E, Himmel DM, Rossmann MG, editors. *International tables for crystallography*. Vol. F, crystallography of biological macromolecules. 2nd ed. New York: Wiley; 2012. p. 333–339.
- [153] Bricogne G. Geometric sources of redundancy in intensity data and their use for phase determination. *Acta Cryst.* 1974;A30:395–405.
- [154] Bricogne G. Methods and programs for direct space exploitation of geometric redundancies. *Acta Cryst.* 1976;A32:832–847.
- [155] Wang BC. Resolution of phase ambiguity in macromolecular crystallography. In: Colowick SP, Kaplan NO, editors. *Methods of enzymology*. Vol. 115. New York: Academic Press; 1985. p. 90–111.
- [156] Zhang KYJ, Cowtan KD, Main P. Density modification and phase combination: phase improvement by iterative density modification. In: Arnold E, Himmel DM, Rossmann MG, editors. *International tables for crystallography*. Vol. F, crystallography of biological macromolecules. 2nd ed. New York: Wiley; 2012. p. 385–400.
- [157] Larson SB, Day J, Canady MA, Greenwood A, McPherson A. Refined structure of desmodium yellow mottle tymovirus at 2.7 Å resolution. *J Mol Biol.* 2000;301:625–642.
- [158] Abad-Zapatero C, Abdel-Meguid SS, Johnson JE, Leslie AG, Rayment I, Rossmann MG, Suck D, Tsukihara T. Structure of southern bean mosaic virus at 2.8 Å resolution. *Nature.* 1980;286:33–39.
- [159] Harrison SC, Olson AJ, Schutt CE, Winkler FK, Bricogne G. Tomato bushy stunt virus at 2.9 Å resolution. *Nature.* 1978;276:368–373.
- [160] Wikoff WR, Liljas L, Duda RL, Tsuruta H, Hendrix RW, Johnson JE. Topologically linked protein rings in the bacteriophage HK97 capsid. *Science.* 2000;289:2129–2133.
- [161] Bragg L, Perutz MF. The structure of hemoglobin. VI. Fourier projections on the 010 plane. *Proc R Soc.* 1954;A225:315–329.

- [162] Blow DM, Crick FHC. The treatment of errors in the isomorphous replacement method. *Acta Cryst.* 1959;12:794–802.
- [163] Unge T, Liljas L, Strandberg B, Vaara I, Kannan KK, Fridborg K, Nordman CE, Lentz PJ. Satellite tobacco necrosis virus structure at 4.0-Å resolution. *Nature.* 1980;285:373–377.
- [164] Carvin D, Islam SA, Sternberg NJE, Blundell TL. Isomorphous replacement: the preparation of heavy-atom derivatives of protein crystals for use in multiple isomorphous replacement and anomalous scattering. In: Arnold E, Himmel DM, Rossmann MG, editors. *International tables for crystallography. Vol. F, crystallography of biological macromolecules.* 2nd ed. New York: Wiley; 2012. p. 317–326.
- [165] Grosse-Kunstleve RW, Brünger AT. A highly automated heavy-atom search procedure for macromolecular structures. *Acta Cryst.* 1999;D55:1568–1577.
- [166] Stubbs MT, Huber R. Isomorphous replacement: locating heavy-atom sites. In: Arnold E, Himmel DM, Rossmann MG, editors. *International tables for crystallography. Vol. F, crystallography of biological macromolecules.* 2nd ed. New York: Wiley; 2012. p. 327–332.
- [167] Terwilliger TC. Automated structure solution, density modification and model building. *Acta Cryst.* 2002;D58:1937–1940.
- [168] Caliendo R, Carrozzini B, Cascarano GL, De Caro C, Giacovazzo C, Monstakimov M. The partial structure with errors: a probabilistic treatment. *Acta Cryst.* 2005;A61:343–349.
- [169] Sheldrick GM, Gillmore CJ, Hauptman HA, Weeks CW, Miller R, Uson I. Direct methods: Ab initio phasing. In: Arnold E, Himmel DM, Rossmann MG, editors. *International tables for crystallography. Vol. F, crystallography of biological macromolecules.* 2nd ed. New York: Wiley; 2012. p. 413–432.
- [170] Xu H, Hauptman HA. Recent advances in direct phasing methods for heavy-atom substructure determination. *Acta Cryst.* 2006;D62:897–900.
- [171] Dickerson RE, Weinzierl JE, Palmer RA. A least-squares refinement method for isomorphous replacement. *Acta Cryst.* 1968;B24:997–1003.
- [172] de la Fortelle E, Bricogne G. Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods. *Methods Enzymol.* 1997;276:472–494.
- [173] Kleywegt GJ. Use of non-crystallographic symmetry in protein structure refinement. *Acta Cryst.* 1996;D52:842–857.
- [174] Rossmann MG, Blow DM. Determination of phases by the conditions of non-crystallographic symmetry. *Acta Cryst.* 1963;16:39–45.
- [175] Rossmann MG, Arnold E. Molecular replacement: noncrystallographic symmetry averaging of electron density for molecular-replacement phase refinement and extension. In: Arnold E, Himmel DM, Rossmann MG, editors. *International tables for crystallography. Vol. F, crystallography of biological macromolecules.* 2nd ed. New York: Wiley; 2012. p. 352–363.
- [176] Brünger AT, Adams PD, Rice LM. Refinement: Enhanced macromolecular refinement by simulated annealing. In: Arnold E, Himmel DM, Rossmann MG, editors. *International tables for crystallography. Vol. F, crystallization of biological macromolecules.* 2nd ed. New York: Wiley; 2012. p. 466–473.
- [177] Ten Eyck LF, Watenpaugh KD. Refinement: introduction to refinement. In: Arnold E, Himmel DM, Rossmann MG, editors. *International tables for crystallography. Vol. F, crystallography of biological macromolecules.* 2nd ed. New York: Wiley; 2012. p. 459–465.
- [178] Brünger A. Refinement of X-ray crystal structures. In: Egelman E, editor. *Comprehensive biophysics. Vol. 1, biophysical techniques for structural characterization of macromolecules.* Amsterdam: Elsevier; 2012. p. 105–115.
- [179] Tronrud DE. Introduction to macromolecular refinement. *Acta Cryst.* 2004;D60:2156–2168.
- [180] Chopra G, Summa CM, Levitt M. Solvent dramatically affects protein structure refinement. *Proc Natl Acad Sci USA.* 2008;105:20239–20244.
- [181] Larson SB, Day J, Greenwood A, McPherson A. Refined structure of satellite tobacco mosaic virus at 1.8 Å resolution. *J Mol Biol.* 1998;277:37–59.
- [182] McPherson A. Virus RNA structure deduced by combining X-ray diffraction and atomic force microscopy. In: Klostermeier D, Hamman C, editors. *Structure and folding of RNA.* Berlin: deGruyter; 2014. p. 125–156.
- [183] Chen ZG, Stauffacher C, Li Y, Schmidt T, Bomu W, Kamer G, Shanks M, Lomonosoff G, Johnson JE. Protein-RNA interactions in an icosahedral virus at 3.0 Å resolution. *Science.* 1989;245:154–159.

- [184] Johnson JE, Rueckert RR. Packaging and release of the viral genome. In: Chiu W, Burnett RM, Garcea R, editors. Structural biology of viruses. Oxford: Oxford University Press; 1997. p. 269–287.
- [185] Tihova M, Dryden KA, Le TV, Harvey SC, Johnson JE, Yeager M, Schneemann A. Nodavirus coat protein imposes dodecahedral RNA structure independent of nucleotide sequence and length. *J Virol.* 2004;78:2897–2905.
- [186] Lucas RW, Larson SB, Canady MA, McPherson A. The structure of tomato aspermy virus by X-ray crystallography. *J Struct Biol.* 2002;139:90–102.
- [187] Kleywegt GJ, Jones TA. Where freedom is given, liberties are taken. *Structure.* 1995;3:535–540.

Subject index

- (GroEL-KMgATP)₁₄ complex, 31
 260/280 absorption ratio, 8
 Aberrant particles, 17
 Absences, 17
 Absorption, 29
 Accuracy, 29, 34, 45
 Additives for crystallization
 Adenovirus, 6
 AIMLESS, 32
 Alfalfa mosaic virus (AMV), 7
 Alpha helix, 35, 38
 Alternate amino acid conformations
 Amino acid, 5, 23
 Amino terminal tail, 39
 Ammonium sulphate, 8, 9
 Angular increment, 27, 28
 Anions, 8
 Anisotropic intensity distribution
 Anisotropic temperature factors, 40
 Anomalous pairs, 29
 Anomalous particles, 19
 Anomalous reflections, 28
 Apple core, 27
 Asymmetric unit, 21
 Atomic force microscopy (AFM), 12
 Averaging, 22
 Background, 28
 Background intensity, 24
 Bacteriophage, 14
 Bacteriophage tails
 Batch method
 Beam centre, 26, 28
 Beam intensity fluctuations, 29
 Birefringence, 21
 Black beetle virus, 6
 Blow–Crick refinement, 37
 Blue tongue virus, 6
 Bond dictionary
 Boot strap, 34
 Bragg angle
 Bragg scattering, 19
 Brome(grass) mosaic virus (BMV), 6
 Bulk solvent correction, 42
 Bulk water, 41
 Capillaries, 13
 CCD detectors, 25
 Centrifugation, 7
 Chemical environment, 40
 Chiral asymmetric unit, 22
 Clone, 6
 Coat protein, 6, 14, 35
 Collimation, 25
 Completeness, 29, 30
 Conformational classes, 23
 Constrained refinement, 39
 Correlation among intensities from NCS
 Correlation coefficient of random half sets (CC_{1/2}), 32
 Cowpea chlorotic mottle virus (CCMV), 34
 Cowpea mottle virus (CPMV), 7
 Cocksfoot mottle virus (CfMV), 35
 Cricket paralysis virus, 6
 Cryo-cooling, 12
 Cryo-electron microscopy, 23, 35
 Cryopreservation, 11
 Crystal decay, 24
 Crystal to detector distance, 25, 27
 Crystallization, 4, 6–11, 14, 17
 Crystallographic unit cell
 CsCl gradients, 7
 Cubic solids, 20, 21
 Cucumber mosaic virus (CMV), 15
 Cultured cells
 D*Trek, 28
 Data collection, 1, 12–14, 22, 24–29, 31, 35
 Data collection strategy, 26
 Data management, 22
 Data processing, 1, 27–31
 Data quality, 29–31
 Defect density, 19
 Defect formation, 14
 Defects, 12, 13, 17–19, 29
 Dehydration, 14
 Density modification, 1, 37, 45
 Desmodium yellow mottle virus (DYMV), 34
 Detector, 25–27
 Detergents, 8
 Developing surface lattice, 15
 Difference electron density map, 32
 Difference Fourier synthesis
 Diffraction limit, 24, 26, 29, 31
 Divalent cations, 8
 Divergence, 25, 28
 DNA, 4, 5, 7, 8, 41, 43
 Domain, 12, 19, 20
 Domain boundary (boundaries)
 Domain motions, 32
 Double helical RNA, 43
 Dodecahedra, 20
 Dust particles, 16
 Effective resolution, 29, 32

- Elasticity, 19
- Electron density averaging, 1, 34, 37
- Electron density map, 29, 37, 38, 41, 45, 46
- Empty capsids, 7
- Empty virions, 43
- Encapsidate, 4, 7
- Envelope, 13, 14, 28, 35, 37–39
- Equivalent reflections, 24, 33
- Error estimate
- Ethylene glycol, 12
- Exposure, 6, 12–14, 24–27
- Expression, 7
- Extinction, 21
- Face normal, 15
- Face normal growth, 16
- Fibre diffraction, 4
- Fibre loop, 12, 13, 28, 29
- Fibres, 16
- Filamentous viruses
- Fine slicing, 27
- Flash-cooling, 12, 14, 28
- Foot and mouth disease virus, 6
- Foreign particles, 12
- Free interface diffusion, 24
- Freezing damage, 24
- Friedel reflections
- Genomic nucleic acid, 4
- Glycerol, 12
- Goniostat, 26
- Group II intron, 32
- Growth kinetics, 14
- Growth mechanisms
- Growth steps, 15, 16
- Hanging drop
- Helical viruses
- Helper viruses, 7
- Hexanediol, 9
- High-resolution limit, 29–32
- Highest resolution shells, 29, 30, 32
- HK-97, 4, 13
- HKL2000, 28
- Homology, 34
- Host cell enzymes, 5
- Host cells, 7
- Host plants, 6
- Human immunodeficiency virus, 7
- Human rhinovirus, 10
- I/σ_1 , 29–32
- Ice, 12, 13, 28
- Icosahedral axis
- Icosahedral symmetry, 1, 23, 30, 36, 40–42
- Ideal geometry, 29
- Image plates, 25
- Impurities, 16, 18
- Independent reflections, 13, 24, 39
- Indexing, 26, 28
- Initial crystallization conditions
- Initial phases, 1, 35, 36
- Insect cells, 5, 6
- Insect viruses, 5
- Integration of intensity
- Inter-particle contacts, 40
- Interstitial spaces, 12, 17
- Intersubunit contacts, 39
- Ions, 8, 35, 36, 41, 43
- Iridovirus, 6
- Isomorphous heavy atoms, 34
- Jelly roll beta barrel
- Lattice interactions, 14
- Lattice strain, 12
- Lattice vacancies, 12
- Line defects, 18, 19
- Liquid nitrogen, 12
- Local lesions, 6
- Localized disorder, 12
- Low completion in high-resolution shells
- Low ionic strength buffers, 10
- Low-resolution X-ray reflections, 35
- Lysozyme crystal, 26
- Macromolecules, 5, 8, 14, 22
- Map averaging, 30, 37
- Mason-Pfizer monkey virus, 7
- Maximum likelihood, 30, 31, 37, 44
- Measurement redundancy
- Merging, 14, 29–32
- Methylpentanediol (MPD), 8
- Micro-beams, 11
- Microcrystals, 16–18
- Microdialysis, 8
- Microorganisms, 5, 6
- Misoriented microcrystals, 12, 16
- Mixed infection, 6
- Model bias, 37
- Model influence, 44
- Model quality, 29, 30, 32, 45
- Model statistics, 29
- Molecular replacement, 34, 35
- Molecular substitution, 1, 34
- Mosaic blocks, 19–20
- Mosaic character, 12
- Mosaicity, 12, 24, 28, 29
- MOSFLM, 28
- Mother liquor, 9, 11, 12, 15–18
- Multi-partite genomes
- Multiple crystal forms, 11
- Multiwire detectors, 25
- Mutant viruses
- Nanolitre volumes, 11
- Nominal resolution, 30
- Non-crystallographic symmetry (NCS), 13
- Non-ionic detergents, 8
- Nucleation, 14–16
- Nucleic acid, 4, 7–8, 22, 39, 41, 42–44
- Nucleic acid helix, 43
- Nudaurelia capensis omega virus
- Observation to parameter ratio, 39, 40
- Optical axis, 21
- Optical effects, 21
- Oscillation image, 25, 26
- Outliers, 29, 30
- Overlapping reflections, 28, 29
- Packing considerations, 34
- Panicum mosaic virus (PMV), 7
- Parasite crystal, 28
- Partial model
- Partial reflection
- Particle deformation
- Particles per unit cell (Z), 24
- Patterson technique, 36
- pH, 8–11
- Phase determination, 1, 34–35, 37, 39, 45
- Phase extension, 34–36, 38, 44, 45

- Phasing, 22, 23, 34, 37
- Planar defects, 12, 19, 20
- Plant viruses, 5, 6, 35, 38
- Plasmid, 6
- Platonic solids, 20
- Point defects, 18, 19
- Point group, 3, 5, 14
- Polarized light, 21
- Poliovirus, 23
- Polyethylene glycol (PEG), 7
- Polymorphs, 10
- Precipitation, 7
- Precision, 29–32, 34, 38, 39
- Pre-nucleation, 14
- Preservation, 35
- Probe model, 34–37
- Protein Data Bank, 8
- Pseudosymmetry, 33, 44
- Purification ‘tags’, 7
- Quasi-elastic light scattering (QELS), 14
- Quasi-equivalence, 23
- Quasi-symmetry, 14
- R* factor, 31–34, 37, 44
- Radiation damage, 12, 13, 24, 26, 27
- Ramachandran plots, 45
- Randomly oriented crystals, 27
- Real space, 44, 46
- Reassembly
- Reciprocal lattice spacings, 24, 26, 29
- Reciprocal space, 27, 28, 33, 38, 44, 45
- Recombinant DNA techniques, 7
- Redundancy independent, 31
- Redundant, redundancy
- Refinement, 3, 4, 29–32, 34, 37–45
- Reflection centre, 26
- Reflection divergence
- Reflection overlap, 25
- Refractive index, 21
- Resolution, 3, 15, 18, 20, 21, 24–26, 29–34, 36–41, 45, 46
- Resolution shells, 29–32, 45
- Restrained refinement, 40
- R_{free} , 29, 32, 44, 45
- R_{meas} or R_{rim} , 31
- R_{merge} , 29–32
- RNA, 4, 5, 7, 8, 41–44
- Robotic methods, 11
- Rod-shaped viruses, 4
- Rotating anode sources, 25
- Rotation function, 33, 34, 44
- Rotavirus, 6
- R_{pim} , 31
- R_{work} , 32
- Salt fractionation, 7
- Satellite panicum mosaic virus (SPMV), 7
- Satellite tobacco mosaic virus (STMV), 6
- Satellite tobacco necrosis virus (STNV), 6
- Satellite viruses, 7
- Saturation
- SCALA, 32
- Scaling, 12–14, 29, 33
- Scaling statistics, 33
- Screw axis, 24
- Screw dislocations, 12
- Self-assembly
- Sequential layer addition, 15
- Sesbania mosaic virus, 10
- Signal-to-noise ratio, 31, 32
- Sitting drop, 8
- Slippage, 28
- Small-angle diffraction
- Solvent content, 10, 12, 19
- Solvent flattening, 34, 37, 38, 46
- Space group, 14, 21, 22, 24, 27, 33
- Special positions, 14, 33
- Spherical viruses, 4
- Spiral dislocations, 15, 19
- Spot separation, 25, 28
- Spot size, 25
- ss-RNA, 4
- Stacking faults, 12, 19
- Statistical residuals, 45
- Step bunches, 16
- Step edge movement, 20
- Step edges, 15–17, 19
- Structure amplitude test set for *R* factor
- Structure amplitudes, 3, 27, 29, 30, 33, 34, 37, 44, 46
- Structure factor difference Fourier synthesis, 44
- Structure solution, 23, 28, 29, 31, 32, 34
- Structure-specific nucleic acid binding sites, 42
- Supersaturation, 8, 15
- Swollen viruses
- Symmetry, 3–6, 8, 10, 11, 13, 14, 20–24, 26, 28–30, 32–34, 36, 37, 39–44
- Symmetry equivalent, 25, 29, 30, 33
- Synchrotron, 11–14, 23–25
- Systematic error, 29
- Systemic infection
- Temperature factors, 39–43
- Terraces, 15, 19
- Thermodynamic and kinetic parameters for crystal growth
- Three-dimensional nucleation, 16
- Tobacco mosaic virus (TMV), 7
- Triangulation number, 7, 14, 23
- Turnip yellow mosaic virus (TYMV), 6
- Twin, 24
- Twinning, 28
- Two-dimensional nucleus (nucleation)
- Ty3 retrotransposon, 7
- Ultraviolet light
- Vacancies, 12, 13, 17–19
- Vapour diffusion, 8
- Vectors, 7
- VIPER data base, 34
- Viral genomes, 8
- Virus centre position
- Virus-like particles (VLPs), 7
- Virus purification kits, 7
- Virus resistant strains, 6
- Virus solubility, 10
- Water molecules, 32, 41, 42
- Weak data, 30
- XDS, 28, 32
- X-ray exposure, 24, 31
- X-ray intensity