Annual Reports C

RSC Publishing

REVIEW

View Article Online
View Journal | View Issue

Molecular structure by X-ray diffraction

Harold R. Powell*

DOI: 10.1039/c3pc90004e

X-ray crystallography has matured over the course of the last century to be the method of choice for the determination of solid-state structure. This is reflected in the lack of new developments in methods for solving and refining structures by single crystal techniques, although some advances have been made in understanding the underlying methods and extending applications to address more difficult problems; most of these are in macromolecular crystallography rather than in the small molecule field. The main areas of progress have been in sample preparation, in X-ray source and detector technology, in the software required to process the data obtained, and in applying sophisticated statistical methods to the analysis. A major visible consequence of the continued importance of the science is that the main crystallographic databases continue to grow rapidly, each with similar doubling rates to those found for the last several decades.

Introduction

2012 marks the centenary of the demonstration of X-ray diffraction by single crystals in the laboratory of Max von Laue by his co-worker Walter Friedrich and his student Paul Knipping in Göttingen.[1,2] Von Laue's breakthrough was his realisation that the interatomic distance in crystals was similar to the wavelength of X-rays, and that if crystals were regular arrays of atoms (or molecules), then a single crystal would behave in the same way to X-rays as a diffraction grating does to visible light. By the end of that year, the first report appeared of a crystal structure determination (that of zinc blende) by means of X-ray diffraction by William Lawrence Bragg (*i.e.* the son of William Henry Bragg).[3] Von Laue was awarded the Nobel Prize in Physics in 1914 for his work in this field, while the Braggs were awarded the Physics Prize jointly in 1915; W. L. Bragg remains the youngest Nobel laureate (at 25) to this day. W. L. Bragg also realised that the interference pattern produced by X-ray diffraction from a crystal could be visualised as arising from reflection of the X-rays from planes of lattice points; this simplified interpretation has now been taught to generations of students as ''Bragg's Law''.

The work of von Laue and the Braggs led to a complete change in the science of Crystallography; where previously it had been the scientific equivalent of

MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge Biomedical Campus, Cambridge CB2 0QH, UK. E-mail: harry@mrc-lmb.cam.ac.uk

stamp collecting, it developed into the gold standard for analysing the atomic and molecular structure of hard condensed matter. By the early 1920's the method had been extended to organic chemicals (*e.g.* hexamethylene-tetramine[4]), and although the first high quality diffraction photographs of protein crystals had been obtained by 1934 (of pepsin, by J. Desmond Bernal and Dorothy Crowfoot (later Hodgkin)[5]), the first protein structure determined by X-ray methods was not reported until 1958 – that of myoglobin (an oxygen storage protein found in muscles) by John Kendrew and co-workers.[6] Although the structure of the DNA double helix was determined by X-ray methods in 1953, it should be noted that the sample was in the form of extended wet fibres rather than as three-dimensional crystals;[7] the first true *crystal* structure of DNA was published in 1979.[8]

To mark the centenary of the discovery of X-ray diffraction from crystals, the International Union of Crystallography (*IUCr*) published special editions of *Acta Crystallographica Section A*. The January 2012 issue (published jointly with *Zeitschrift für Kristallographie*) concentrated on the Laue centennial and was based on lectures given at the 2012 meeting of the Deutsche Gesellschaft für Kristallographie,[9] while the January 2013 issue focussed on the Braggs and their discoveries, and was collected from talks at the Bragg Centennial Symposium held in Adelaide in December 2012.[10] Taken together, they provide a potted history of Crystallography from the 18th century onwards. Both editions are available as open-access (*i.e.* free to the reader) PDF downloads from the *IUCr*.[11]

1 Crystal growth

Growing protein crystals in a microgravity environment has been shown previously to be advantageous under some specific circumstances, but only on dedicated space flight missions, and was also dependent on the number of times a particular protein had been tried.[12] The two main effects on crystallisation that are reduced to negligible levels in microgravity are convection effects and sedimentation, which can also be reduced on Earth by performing the crystallisation in a highly viscous medium such as a gel; working in a terrestrial environment has huge advantages in reduced cost and ready access to the experiment. Over the last few years, microfluidic devices have been developed for many applications in chemistry and biology, and their use in crystallisation is a natural further development. Previously they had only been used in screening studies to optimise conditions rather than to grow crystals suitable for diffraction experiments, but this has now been tried.[13] The ratio of the buoyancy of a crystal to the viscous force acting on the crystal ("buoyancy convection") may be described by the dimensionless Grashof number Gr_D, which is proportional to gL^3/ν^2, where g is the acceleration due to gravity, L is the characteristic length (which can be thought of as the height or width of the vessel containing the solution) and ν is the kinematic viscosity. It is apparent that decreasing L or increasing ν would give rise to analogous effects to decreasing g. A glass microfluidic device with a 100 μm circular crystal growth microchannel has been used to grow crystals of hen egg-white lysozyme (HEWL), by evaporation of the solvent through the membrane. HEWL is probably the most common test protein for early-stage development of protein crystallographic methods. Calculating Gr_D shows that changing a microchannel length to ~93 μm from 2 mm is equivalent to reducing g to around 10^{-4} times its value at the Earth's surface. Since convection

and sedimentation effects are all but eliminated, the principle means of mass transport of protein to the crystal surface is diffusion. It was noted that the size of air bubbles in the microchannels was important in determining the rate of evaporation; larger bubbles were associated with faster evaporation, and the size of the bubbles could be controlled to some extent by modifying the relative humidity around the microfluidic device. High quality, large ($\sim$1 mm) HEWL crystals could be grown before protein depletion in the solution began to affect crystal growth. Once the mass transport of protein to the crystal surface becomes slower than the surface kinetics of crystal growth, crystals tend to develop irregular surfaces (or are "morphologically unstable").

In contrast to HEWL, membrane proteins are extremely difficult to crystallise, because the molecules have two hydrophilic regions at either end (*in vivo* these reside on either side of a lipid bilayer) and a central lipophilic "belt" that sits within the membrane. A well-established technique for growing crystals of these species for diffraction studies is to crystallise from a cubic lipidic mesophase (CLM), which consists of a three-dimensional infinite membrane phase interpenetrated by a 3D continuous aqueous channel system.[14] This medium is difficult to handle due to its stickiness and high viscosity, so in spite of its success in producing membrane protein crystals, it is used in remarkably few laboratories worldwide. A protocol designed for learning to work with CLM, that uses the globular protein HEWL has been published.[15] Becoming adept in handling the reagents in conjunction with a cheap and readily available protein means that the experimentalist will have extra confidence when setting up crystallisation trials with their hard-to-obtain membrane protein. It was noted that the environment around the crystals bears some similarity to that found in gels or in microgravity.

The goal in most crystallisations in structural science is to obtain perfect single crystals, but the analysis of crystal perfection is usually ignored. Topographic studies of single crystals (*i.e.* analysing the shape of Bragg peaks) using collimated X-ray beams have been reported using X-ray film as the detector; even though X-ray film was made effectively obsolete for structural diffraction studies in the 1990's, it is still the detector of choice in topography. The main advantage of CCDs over film is that of speed, as many images can be obtained and made available for subsequent processing in the time taken for a single film to be exposed, developed and fixed. The seeming reluctance to adopt electronic detectors seems to be mainly due to the non-availability of electronic detectors for X-rays with sufficiently small pixels to be able to map dislocations and other imperfections in the samples under examination. Typically, X-ray film can be scanned easily to give images with pixel edge 10–15 μm, though the limiting spatial resolution is much lower. A CCD detector with 6.45 $\times$ 6.45 μm pixels (in 3 $\times$ 3 binned mode, giving an effective pixel size of $\sim$20 $\times$ 20 μm) has been used in a study of large ($\sim$2 mm) HEWL crystals in order to determine its suitability in this application.[16] In this work, the exposure times for X-ray film were 600 s, compared to 0.5 s for the CCD data, and while the images obtained from the CCD are of lower resolution, dislocations could be readily identified.

2 Sources and detectors

X-ray free electron lasers (XFELs), which produce extremely short intense pulses offer a new light source for crystallographers. The pulses produced by the Linac

Coherent Light Source in Stanford repeat at a rate of ~120 Hz, while the pulses themselves have a lifetime of the order of femtoseconds. The crystalline sample is destroyed in a single pulse, but remains integral during the passage of the photons through it; if a new crystalline sample can be introduced into the X-ray beam for every pulse, it is possible to collect many images which may be able to be built up into a complete dataset in a process that has become known as ''serial femtosecond crystallography''. This may be achieved by suspending microcrystals in a liquid carrier, and directing a jet of this suspension through the path of the incident X-rays. The drops are formed by the ''electrospray'' method; if a sufficiently high electric field is present on the jet, surface tension can be overcome and the free surface becomes focussed into a microjet which breaks apart into a regular series of microdrops.[17] It is possible to engineer the concentration of microcrystals so that, on average, a single crystal is present in each microdrop. Additives such as glycerol can be used to modify the length of the microjet before it breaks up.

It is likely that the crystal sample in the suspension will be more dense than the suspending liquid, in which case it follows that settling may be a problem; this has been observed to occur over several minutes, even with extremely small crystals. In order to combat this, an ''anti-settling'' device has been developed for use with the small total volumes involved (typically <5–6 ml). The device rotates the sample chamber back and forth about a horizontal axis through a range of 350° (to avoid tangling of pipework and electrical cabling) at a rate of 0–90° min^{-1}.[18] An initial design using a commercially available syringe pump was developed to operate at a pump pressure of up to 6 MPa, but was not suitable for more viscous media such as polyethylene glycol (PEG) solutions. Since PEG is a common component in protein crystallisation protocols, a high pressure version was developed subsequently using an HPLC pump capable of delivering a pressure up to 40 MPa.

Processing diffraction data in serial femtosecond crystallography presents a number of problems not encountered in a ''normal'' data collection. 120 diffraction images can be collected every second; if there is one crystal per microdrop, each image will contain the diffraction pattern of a single crystal in an arbitrary orientation (assuming there is no preferred orientation in the microjet). As the X-ray pulse is only a few tens of femtoseconds, the orientation of individual crystals will not change during the course of the exposure, and consequently all Bragg peaks will be partially recorded. In addition, when the crystals are very small, they consist of only a few unit cells, and this is reflected in the shape transform of the crystal recorded on the detector as a series of interference fringes associated with every diffraction maximum. To deal with these problems, a software suite called *CrystFEL* has been developed.[19] *CrystFEL* is written largely in *C*, and its components are ''glued'' together with *Perl* and *UNIX* shell scripts. The autoindexing and integration of each diffraction pattern is performed by a script (''*indexamajig*'') that utilises existing programs for indexing (*DirAX*[20] and *MOSFLM*[21]) and distributes the task among many processors. Following successful indexing, the images are integrated using a simple summation method, *i.e.* the pixel counts in each spot are added together and the estimated background level is subtracted. Profile fitting the reflections is not used because the peak shapes were found to vary widely. Intensity measurements from different images of reflections with the same indices are merged together using a Monte Carlo method (in the script ''*process_hkl*'') to give estimates of intensities of the fully

recorded reflections. It was noted that if merohedral twinning is possible for the space group used for this processing, then all intensity measurements produced will correspond to those from a perfect twin. *CrystFEL* has been used to process datasets collected from microcrystals of samples ranging from the archetypal test protein HEWL to the membrane protein complex Photosystem I.[22] In addition to the raw processing aspect of *CrystFEL*, the suite provides utilities for visualising the data and performing further analysis of the processed images and intensities.

The intense radiation available at XFELs has long been promoted as being ideal for so-called ''single molecule X-ray diffraction''; this represents the limiting case for diffraction studies, since the ''crystal lattice'' consists of a single asymmetric unit.[23] In contrast to serial femtosecond crystallography, a single molecule interacts with each flash of X-rays rather than a single crystal. For molecules much smaller than large viruses, the quantum noise associated with the extremely weak signal (remembering that the crystalline lattice effectively amplifies the signal) remains a significant hurdle to the method. A two step algorithm has been developed which attempts to combine the two-dimensional scattering patterns recorded from many thousands of molecules into a single three-dimensional intensity density function of the scattering, and in the process determine the limiting resolution of the method.[24] Each molecule is in a random orientation, so two images may result either from two molecules in the same orientation (in which case the images should be identical, disregarding random variations but allowing for rotation about the vector describing the incident radiation) or in different orientations (where the two images will share only a common circle of intensities). The first step in the process, therefore, is to group images together corresponding to similar orientations in order to reduce the noise, and the second step is to identify the mutual orientations of the groups of images. The first stage is performed by calculating a correlation function between patterns, then normalising the data on concentric circles on each pattern to its mean (which suppresses the quantum noise); if the two patterns correspond to similar orientations, a radial line of high intensity can be observed. Once these groups of patterns have been identified, applying the correlation function between two averaged groups shows either a circle or an arc, corresponding to the intersecting circle (or arc) of data. Using this approach, it is possible to detect a signal where as few as 0.1 photons per ''effective pixel'' have been recorded, and the resolution can be determined as corresponding to the wavenumber of the pixels where the mean value of the correlation function disappears into the background.

While much current attention is directed towards these new high intensity sources, the majority of protein crystallography data collections are performed either in the home laboratory or at a synchrotron. The facilities for macromolecular structure solution at the German synchrotron based in Berlin, *BESSY*, have been described.[25] The report is most useful in that it describes a typical facility, equipped with instrumentation and software commonly installed for data collection and processing at most synchrotron sites around the world. Attention is drawn to the inadequacy of teaching of methods in macromolecular crystallography in most university departments around the world; most new students entering the field come from the life sciences and have not studied X-ray diffraction at all. Many synchrotron laboratories hold courses to remedy this situation for postgraduate students and post-doctoral workers, and that offered

by *BESSY* and its partner institution the Freie Universität Berlin is outlined. A tutorial script with examples of raw data for analysis is freely available for download from the web-page http://www.helmholtz-berlin.de/bessy-mx.

3 Data collection and processing

Hybrid pixel array detectors based on direct detection of X-ray photons by silicon sensors have become common during the last few years at macromolecular beamlines at synchrotrons; they have numerous advantages over other types of detector such as those based on CCDs, which have dominated the area detector market for nearly two decades. These advantages include markedly larger dynamic range, lower noise and higher frame rates. Unlike CCDs, which effectively count only $\sim$10% of incident photons, they can record essentially all photons, so they give a very precise measure of each interaction event. However, each photon takes a finite time to be recorded, and is followed by a short dead time. When the interval between the arrival of each photon (the inverse of the count rate) approaches this dead time, then the efficiency of recording is impaired. For the Dectris Pilatus 6M detector, for example, the error becomes important above $\sim 10^6$ photons per second; for macromolecular crystallography, this is not an issue at present, since the detector writes frames at $\sim$25 Hz, and the maximum counts recorded per pixel are likely to be tens rather than hundreds of thousands. With faster frame rates, or higher X-ray fluxes, errors will begin to become important. Provided the incoming rate is not too high, correction factors can be applied to the recorded values, but these depend on the gain setting of the pre-amplifier, the energy threshold, and also on the time structure of the radiation produced in the synchrotron. The correction cannot, therefore, be a single number, but is best calculated *via* a simulation of the instrumental parameters,[26] which can be calibrated by comparison with experimental findings. Increasing the gain also increases the dead time, while applying a higher energy threshold for detection decreases the dead time. What is more striking is that the dead time is strongly dependent on the ''bunch structure'' of the synchrotron storage ring, *i.e.* how many bunches of electrons are circulating in the ring and the interval between them. The relationship between the observed and true rates has a somewhat complicated dependency on the bunch structure.

The Pilatus is not the only hybrid pixel detector; a new detector is in the process of being developed at the French National Synchrotron, Soleil. The prototype has 6 rows of 3 detector modules arranged in the vertical plane, with each module tilted at a small angle to the vertical; much like the prototype Pilatus 1M,[27] this gives the detector some of the appearance of a tiled roof. The corrections necessary to bring real pixel co-ordinates onto a planar virtual detector are described,[28] together with procedures for identifying dead and ''hot'' pixels, rows of pixels which are masked by adjacent modules, and flat-field and background corrections. It is noted, but not explained, that the pixels on the edges of each module are 2.5$\times$ the size (in both height and width) of the pixels in the main body of each module; this is presumably due to the manufacturing process. To date, the detector has been used for determining the elastic strain and strain heterogeneities in deformed thin metallic films, and also for studying the growth mechanism of multi-walled carbon nanotubes.

Most of the detectors on synchrotron beamlines are still CCD based. An analysis of the point spread function (PSF) has been performed.[29] The PSF may be defined as the spatial distribution of the recorded signal when the detector is illuminated by an infinitesimally thin beam of light. The PSF arises from the construction of CCD detectors; an X-ray sensitive phosphor layer converts the incident photons to visible photons, which are carried to the CCD chip itself through a fibre-optic taper. Since the visible photons are emitted uniformly in random directions, an X-ray beam striking a single sub-pixel point at the front of the detector will be recorded not only at the pixel corresponding to the interaction, but also by adjacent pixels as well. By illuminating a single pixel on an ADSC Q315r detector with a 15 μm diameter X-ray beam, and repeating the exposure ~200 times, it was possible to determine that the "tails" of the PSF are due to the fibre optic bundle rather than the phosphor, and that the mathematical form of the PSF follows a Moffat rather than Gaussian or Lorentzian form. The Moffat function may be thought of as a Gaussian convolved with a power law, but may be approximated more simply as a sum of Gaussians. In physical terms, it was found that the PSF resembled the solid angle subtended by a point source 27 μm above the pixel plane. The "solid angle" PSF has been implemented in the diffraction image simulation program *MLFSOM*,[30] and produces more realistic spot shapes than those synthesised using a Gaussian PSF.

The design of X-ray data collection experiments has developed considerably during the last decade or so. Earlier work involved automated analysis of reference diffraction images with the intention of producing a sensible data collection strategy, and much was learned about how to formalise the process; this led to the development of more flexible protocols but this took an inordinate amount of time.[31,32] The advent of ever faster data collection and more reliable and more complex hardware at the beamline (*e.g.* sample mounting robots, mini-kappa goniostats, *etc.*) has increased the need for being able to develop flexible and reliable protocols quickly. A particular shortcoming became evident in these early efforts; if any one part of the process failed, then all subsequent analysis stalled. It was with this in mind that a *Data Analysis Workbench* (*DAWB*) has been developed at the ESRF in Grenoble to implement the idea of "workflows" for development of complex automated protocols which consist of different tasks in data collection and analysis.[33] The idea is based around individual actors being able to communicate with other actors *via* a standard protocol. Each "actor" is the software that performs some specific task, *e.g.* crystal alignment or characterising a crystal, and the structure of the workflow is flexible enough to allow easy inclusion of new or modified actors should any step fail. The use of workflows separates the machine control aspect from the experimental protocols, which in turn means that the experts in either field can concentrate on their own speciality. Since each is working in their own area of expertise, development cycles become much shorter and more ambitious projects can be tackled. An additional advantage is that the experimental protocols can be tested off-line and optimised before being used to interact with physical machines.

It has been known for many years that fine phi sliced data collection (where the oscillation angle for each image is less than the crystal's reflecting range) should, in principle, give better data than coarse slicing. Fine slicing reduces the background associated with individual measurements and also means that the maximum counts per pixel will be lower, reducing the number of saturated pixels,

and for non-overloaded extremely strong reflections a count-rate correction is less likely to be necessary. Exactly how fine the ''slices'' should be depends on a number of different criteria, of which the synchronisation of X-ray shutter with the crystal rotation is particularly important, as is the accuracy and reproducibility of opening and closing the shutter. The advent of fast readout detectors such as the Pilatus has meant that ''shutterless'' data collection can be performed (in which the shutter is kept open throughout the experiment and not closed at the end of each image), although a small proportion of the data is lost between images due to the finite readout time of the detector. The optimal rotation range for each image has been studied for a number of protein crystals.[34] While the quality of the data produced is better for fine phi slicing than coarse slicing (as expected), the optimum value appears to be around half of the mosaicity of the crystal. Conversely, collecting data with too fine a slice means that beamline phenomena with the same time scale as the exposure (for example incident beam intensity fluctuations or crystal vibration) can interfere and reduce data quality. The definition of the mosaicity used in this experiment was the FWHM of a Gaussian rocking curve (*e.g.* as used in programs like *XDS*;[35] other integration programs use definitions which give rise to different numeric values).

The ultra-low temperature *XIPHOS* diffraction apparatus holds a sample which can be cooled to 1.9 K. The chamber is held under vacuum inside a beryllium shield, and, in order to maintain the extremely low temperature, this is located in a second Be container which acts as a thermal radiation shield. Although Be is mostly transparent to X-rays, there is nonetheless significant non-homogeneous powder diffraction from these shrouds (because of randomly distributed large crystallites of Be in the shroud material) which interferes with measuring the diffraction from the sample of interest. For weak samples, the Be diffraction rings are of the same order of magnitude as the signal from the sample, so steps must be taken to remove their effects. A program *Masquerade* has been written which performs this task.[36] Since the beryllium scattering originates at a different distance to the detector than that of the sample, the diffraction from the two origins will move relative to each other if the crystal to detector distance is changed. It is therefore possible to determine the scattering which is due to the shroud and to create masks from this which can then be used to avoid measuring ''contaminated'' reflections. The analysis is complicated because diffraction from the two shields does not have a common origin, and so individual sets of masks must be generated for each. Elimination of reflections in the masked regions by this method gave rise to significantly better refinements of two un-named test structures with data collected at 123 K and also with data collected at 2 K from the organic compound cytidine.

In recent years, manufacturers have developed crystallisation plates which interfere less with X-rays. An integrated stand-alone scanner for plates with a sealed-tube source has been marketed, which allows the user to screen the contents of many sample wells automatically by X-ray diffraction without having to mount individual crystals. This device uses crystallisation plates placed horizontally on it, so that the drops do not slip off to one side.[37] This can be extended to collecting full datasets ''*in situ*'' to allow the crystallographer to collect diffraction data at ambient temperature from crystals while they are still sitting in the crystallisation plate. Among the reasons for doing this include minimising the handling of crystals (protein crystals are fragile, and those of

membrane protein crystals especially so), the study of virus crystals (which are often not amenable to cryocooling, and may be pathogenic, in which case they are best left undisturbed in the crystallisation plate) and the opportunity to screen many samples rapidly in different crystallisation trial conditions before scaling up to produce larger crystals for dedicated diffraction studies.

A high-precision goniostat which can grip crystallisation trays has been developed and mounted on the microfocus beamline I24 at the Diamond Light Source. It includes an accurate XYZ translation stage so that individual microcrystals (around the same dimensions as the X-ray beam, $\sim$10 μm) can be centred accurately in the incident X-ray beam.[38] The problems of optically centring the crystal (the plate material refracts visible light significantly) were addressed partly through calibration with the particular trays used. Also, orienting the trays so that they are vertical means that there is a chance that crystals, especially those suspended in drops, have a chance of slipping downwards.

A simple jig produced for the same purpose for use in the home laboratory can be mounted on an standard X-ray diffractometer.[39] Coarse adjustment is performed *via* a simple scale derived from the plate dimensions, while crystals can be centred accurately by micrometer adjustment screws. This simple device allows several dozen crystals to be examined in a few hours.

There has been a long-term trend to use ever smaller crystals for X-ray data collection, especially in protein crystallography, as X-ray sources have become more powerful and detectors have become more sensitive and less noisy; a typical isometric crystal used nowadays would be no more than about 100 μm across, and often smaller. Crystal mounts such as cryoloops or meshes are often of a similar size, and present different profiles in different projections, so that the background scatter from the mount itself is significant and not isomorphous. For many years, crystals in small molecule data collections have been mounted at the end of glass or silica fibres, so that X-ray scattering from the mount is minimised, since the X-rays do not actually encounter the mount; these hard materials can damage fragile protein crystals, so polyester fibres from the bristles of a toothbrush have been investigated as an alternative.[40] These fibres are drawn (in manufacture) to a width at the tip of $\sim$50 μm, so the contact area between the tip and the crystal is small. Depending on the size of the crystal, a choice can be made as to how many fibres are used to provide sufficient stability, *e.g.* a single fibre is used for crystals in the range 10–50 μm, two fibres for crystals 50–200 μm, and three fibres for those that are larger. The fibres can be mounted in standard magnetic mounts that can be used for automated transfer to a goniostat on a diffractometer. Removing the bulk of the scattering from the mount allows more accurate measurement of data; Bijvoet differences of $<$0.5% were measured in trials and allowed structure solution from the anomalous scattering of sulfurs in native proteins with data collected using Cu-Kα radiation.

Very small crystals are hard to manipulate; those measuring less than 100 μm are difficult to mount, and as noted above, the crystal mount itself makes a significant contribution to the scatter recorded. Before the advent of cryocrystallography, crystals measuring $\sim$200–500 μm were used and were mounted in fluoroborate glass capillaries. The transfer of heat from the crystal *via* the capillary was too slow for the buffer solution to vitrify in flash cooling; crystalline ice would be formed and crystal degradation resulted. A finely drawn capillary

with roughly the same dimensions as microcrystals has been investigated as a mount.[41] Crystals (measuring ~10 μm) are drawn inside by capillary action, and sit on the inside surface of the capillary, with good thermal contact. The wall thickness of ~1 μm allows heat to transfer quickly from the crystal so that water vitrification occurs rather than crystallisation. X-ray scatter from the capillary is substantially closer to that from air than for the polyimide or nylon loops that are more usually employed.

Cryo-cooling protein crystals for X-ray data collection normally involves the search for and use of a "cryoprotectant", *i.e.* an additive that will prevent the formation of hexagonal or cubic ice in the aqueous fraction of the crystal (protein crystals contain from ~30% to ~90% water by volume, depending on the species under investigation). Typical cryoprotectants are small polyols, though other chemicals (including inorganics) have been used. Since the ice formed is polycrystalline, it will usually give rise to powder diffraction rings, and these obscure diffraction maxima at the same resolution and make integration of the spots of interest more difficult. At elevated pressures, *e.g.* around 200 MPa, formation of vitreous (*i.e.* non crystalline) ice is favoured, and this knowledge has been used to develop a cryo-cooling method that avoids cryoprotectants. A crystal growing device which has forty eight internally polyimide coated quartz capillaries with internal diameter 100 μm arranged like the strings of a harp can also be used for *in situ* diffraction, *i.e.* the plate itself can be mounted at a beamline and scanned, so that the crystals grown in the capillaries do not have to be manipulated. Individual capillaries containing protein crystals can be removed from the CrystalHarp, and treated to avoid crystalline ice formation in the following manner; they are immersed in 2-methylpentane (to prevent the formation of air bubbles in the capillaries), pressurised within 6 ms to >200 MPa, held for 290 ms and cooled to <133 K within 30 ms. The temperature is adjusted to at least 108 K to ensure that the 2-methylpentane is liquid, the sample removed to liquid nitrogen (77 K), and then the sample can be treated as any other cryocooled crystal at ambient pressure.[42]

Many organisms use elevated levels of the amino acid L-proline to combat environmental stress, *e.g.* to combat freezing, as it helps to prevent the formation of crystalline ice. This property has been exploited previously in a number of laboratory applications such as using high concentrations of proline in samples which have been freeze-dried, and this prompted an investigation into the use of proline as a cryoprotectant in protein crystallography.[43] Tests using a selection of protein crystals (HEWL, xylose isomerase, histidine acid phosphatase and 1-pyrroline-5-carboxylate dehydrogenase) showed that it was effective in the range 2–3 M, *i.e.* at roughly the same concentration as other widely used cryoprotectants. In common with other organic compounds used for this purpose, electron density maps calculated from X-ray data clearly showed cryoprotectant molecules bound to the proteins.

Most of the time crystallographers rely on the data processing software to accurately model the diffraction experiment, and so yield accurate values of I and $\sigma(I)$. It has been noted that the values of $\sigma(I)$ are often underestimated by the integration and scaling programs.[44] In an effort to determine the true uncertainties associated with the beamline and detector rather than variations in the crystal, an experiment in which many images were collected from a crystal in a single orientation was performed. The data were processed using standard protocols and the resulting values for I and $\sigma(I)$ were analysed against those

parameters which are known to vary with time, *e.g.* decay of the synchrotron ring current;[45] this method is in contrast to an earlier study which analysed a large group of full data sets and concluded that the major source of errors in the strong, well measured reflections was in the experimental set-up.[46] The current study is useful in decoupling the random errors (*e.g.* arising from vibration, turbulence, X-ray background noise, *etc.*) from systematic errors which arise, *inter alia,* from sample properties (*e.g.* crystal non-homogeneity), data integration software and incorrect detector calibrations. The time taken to collect the hundred or so images necessary for this analysis is trivial with a modern detector, so this is a feasible calibration step for any synchrotron beamline. The authors note that, while the average intensity of reflections in this experiment decreases with X-ray dose (giving a rough indication of radiation damage), some individual reflections become more intense; this behaviour reflects the local structural changes that occur with radiation damage. They also noted that the error estimates (obtained from the integration and scaling program used) for weak reflections tend to be greater than expected, while those for the strong reflections were underestimated.

This work is echoed in another study.[47] HEWL crystals were grown under near-identical conditions, and the data collected processed with three of the most widely used processing packages (*DENZO/SCALEPACK,*[48] *MOSFLM/SCALA,*[49] *XDS/SCALA*[35]). Careful analysis of the error estimates obtained from the processing programs showed that they corresponded well with true experimental errors determined by a multiple crystal data collection protocol, for all but the strong reflections. For these, the error estimates from all the programs were found to be too low, and that a significant error inflation is required to account for the true variation from crystal to crystal. Other sources of possible error were considered, including differential X-ray absorption by the sample and unit cell variation,[50] but in themselves were not considered to be sufficient to completely explain the differences. It is well known that the crystallographic *R*-factor (*R*1) for well-determined and refined small molecule structures is often close to the merging *R*-factor (particularly R_{merg}, R_{meas} and other similarly derived quantities) from data processing, but for macromolecular structures (where R_{merg} is substantially less than *R*1 in almost all cases) this is usually not the case. The present study suggests that a good part of this difference may arise from the inadequate error estimates calculated in processing; it has been pointed out recently that the asymptotic behaviour and other properties of the two measures differ markedly; for example, at high resolution R_{merg} tends to infinity and *R*1 remains below 0.40 and so these two indices should not be considered comparable.[51,52]

It was noted in passing that the $I/\sigma(I)$ for area detector data is rarely greater than about 30, even in the lowest resolution shell (which also corresponds to the strongest data),[53] though it was suggested that this may be associated with systematic errors in the instrumental set-up. This $I/\sigma(I)$ limit was addressed independently with an analysis of its significance from data processing programs, with the ''optimistic'' assumption that the behaviour of area detectors can be approximated by Poissonian (or counting) statistics.[54] The authors concluded that the limit exists on average for sufficiently large populations although individual reflections may not be bound by it, and also that the value estimated from Poisson statistics gives a lower limit for the experimental uncertainties; any data processing program that yields larger values for the significance is likely to

be underestimating the errors. Data processed with a variety of programs were analysed and the significance of the reflections was found to vary widely, and it was concluded that the distributions of either the I values or the σ values changed during scaling, which was suggested to be ''suspicious''. Omissions in the calculations were amended subsequently by other authors, who in general agreed with the conclusions in the first paper.[55,56] Overall, it was determined that the empirical scale factors used in some scaling programs gave rise to ''very strange distributions of errors'' which could be avoided by employing an error model based on the variance of multiply observed data rather than one based on a refinable instrument error model.

The diffraction data processing program *XDS* has become increasingly popular in recent times although it has existed for several decades;[5] the program only has command-file input, so it requires the user to be familiar with the command syntax before they can process data effectively (or rely on a previous user to supply command files for them). To get round this problem, an interface *XDSAPP* has been written in the scripting language *TclTk*.[57] *XDSAPP* not only sets up a processing run but it also presents graphs of the variations in refinable parameters that are observed during processing. Interestingly, because *XDS* itself does not read image headers (which often contain useful data processing information, such as sample to detector distance or radiation wavelength) *XDSAPP* runs another widely available data processing program, *MOSFLM,* to extract this information so that an appropriate command input file can be produced automatically.[21]

Wilson statistics are based on the unit cell having homogeneous contents. While this may be a good approximation for macromolecular samples, for small molecules this is often not the case, *e.g.* if a heavy atom is present among many lighter atoms, when there are only very few similar atoms present, or when weak reflections are omitted from the dataset. It can be shown that the probability density function (p.d.f.) can also be affected by the space group,[58] for example, the p.d.f. for a compound with asymmetric unit contents $AuSPC_{25}H_{22}$ that crystallises in $P2_12_12_1$ is close to ''ideal'', whereas that of a compound with similar asymmetric unit contents $AuSPNOC_{28}H_{45}$ in space group *Pbca* differs markedly from ideal. The author points out that the statistics themselves are not at fault but that there are structural elements present which violate the *a priori* assumptions made.

The proper treatment of twinning has become routine in protein crystallography, as it has been for decades in small molecule crystallography. It can be detected by analysing the distribution of intensities in a dataset; for example, a hemihedrally twinned crystal will have fewer very strong reflections (and also fewer very weak reflections) than those expected from Wilson statistics, since the observed intensities are given by the fraction-weighted linear combination of the untwinned true intensities from each component domain. This analysis can be obscured by additional non-crystallographic symmetry (NCS) elements such as pseudo-centering, where half of all reflections are weak anyway and so skew the distribution.[59] If a hemihedral twin has an NCS screw axis with 180° rotation coincident with the twin axis and shift of 1/4 along the axis, half of the reflections that should be indicative of twinning will instead have intensities equal to those for untwinned reflections.

4 Radiation damage

Radiation damage affects the structure of the molecules in a crystal. Where there is a heavy atom sub-structure (*e.g.* heavy metals or samples with disulfide bonds), it is likely that this will also be affected. The method of radiation-damage induced phasing (*RIP*) exploits the changes caused in this way; a reference dataset is collected, the crystal is subjected to a high X-ray dose "burn" to cause damage deliberately, then a second dataset is collected to complete the experiment.[60] A major problem has been in determining the optimum "burn" in order to gain the maximum phase information. A method using a single extended sweep of data collection has been developed; a segment of data contained within this range is identified by statistical means as being the "burn" region.[61] This has several advantages over the established method, namely a single data collection is performed, and the decision about which images are used for the "before" and "after" segments can be optimised off-line after the data collection has been completed. Since radiation damage is inevitable after sufficient time, it is usually enough to merely extend the data collection beyond what would normally be done. The authors noted that even though data completeness in both segments was important in successful structure solution, as was as increasing the gap between the segments, in one favourable test case it was possible to solve the structure with low completeness (down to $\sim$50%) and essentially no gap. The total dose absorbed by the crystals in this investigation was, in all cases, substantially lower than the Garman limit of 2×10^7 Gray (1 Gray, Gy = 1 J kg^{-1}), which gives an estimate of the useful lifetime in an X-ray beam of a protein crystal.[62]

Radiation damage to crystals can be thought of as either "site-specific" or "global"; site specific damage is seen in the changes to the molecule(s) at specific localised sites within many copies in the crystal and is seen in electron density maps in real space; changes in the molecular structure affect the overall diffraction pattern (in reciprocal space), so can be viewed as global changes. A systematic study on crystals of the small protein thaumatin has been performed at 300 K and at 260 K over a wide range of radiation dose rates to see if this can be controlled to "outrun" radiation damage, *i.e.* to collect the data before the sample has been seriously damaged.[63] A small dose rate dependence (as estimated from increase in *B*-factor, decrease in average $I/\sigma(I)$, or change in the unit cell parameters) over a range of dose rates from 6.5 to 103 kGy s^{-1} was observed at both temperatures, but this observation was complicated by a large scatter in values between apparently identical crystals. A crude estimate indicated a decrease in sensitivity by a factor of $\sim$1.5 at 300 K and $\sim$2 at 260 K. This may be a useful result; although the dose rate dependence of radiation damage is small at cryogenic temperatures (*e.g.* 100 K, the most common data collection temperature for macromolecules), the structural information obtained may not be as directly relevant to the biological system as that from data obtained near ambient temperatures. Ice formation is unlikely to occur at 260 K provided surface moisture is removed from the sample prior to data collection, although the high concentration of solute (typically $\sim$2–3 M) in the interstitial channels within the crystal provides only about 5° of antifreeze protection. Other factors such as heating of the sample by the X-rays were considered, but it was concluded that these would be small in comparison to radiation damage effects. In order to study specific site directed radiation damage effects, data collections at five temperatures from

25 K to 300 K from crystals of thaumatin and urease were performed, and the results examined.[64] Previous work has shown that some specific amino acid residues (*e.g.* those containing sulfur or carboxylate groups) are particularly badly affected by radiation damage. This study revealed that in the higher temperature ranges (180–300 K) residues in more mobile regions of the protein, more or less regardless of chemical type, were damaged more than those which lacked mobility; in thaumatin these corresponded to solvent-exposed turns, and those in urease to the active site ''flap''. Additionally, these residues were identified as those most likely to move substantially once damaged. Since these residues are on the surface of the protein, and hence are accessible to free radicals in the solvent channels of the crystals, this may not be surprising, but less mobile residues on the surface did not appear to be damaged as gravely. A numerical analysis of the tightness of the packing of individual residues supported this conclusion.

Fast area detectors such as the Dectris Pilatus currently allow data collection rates of up to 25 images per second in the shutterless mode. Modern synchrotron beamlines such as the microfocus facility on station I24 at Diamond Light Source can deliver fluxes of around 1 MGy s^{-1}. Taken together, these present an opportunity to test whether it is possible to ''outrun'' radiation damage by collecting data very quickly while using a high dose rate.[65] Data were collected from crystals of three different proteins (immunoglobulin γ Fc receptor (FcγRIIIa), bovine enterovirus serotype 2 (BEV2), and the A_{2A} adenosine G-protein coupled receptor (A_{2A}AR)) in two modes, either shutterless or ''stop-start'', in which artificial extra delays of a few seconds were inserted between successive images so as to simulate traditional data collection with slower readout detectors. All the samples decayed more-or-less completely within 25 images, so a large number of crystals were used in the experiments to obtain statistically meaningful results. It should be noted that the structures of FcγRIIIa and BEV2 had previously been solved with room temperature data collected at I24, so only having a few images from these samples did not prevent structure solution. It was found that the sample lifetimes increased as a function of dose rate for exposure times less than 60 ms, *i.e.* when collecting data at rates >16 images per second, which corresponded to dose rates of 0.6–1 MGy s^{-1}. The authors identified three possible origins for this dose-rate effect, each of which is concerned with the fate of hydroxyl radicals in the crystal, namely self-combination of radicals, diffusion rate of radicals through the crystal, and quenching of radicals by bulk solvent. In combination with *in situ* data collection, this appears to be an attractive option that allows structure solution from room temperature data.

Radiation damage is exacerbated by the use of X-radiation in the normal energy ranges favoured by crystallographers, typically 11–14 keV (corresponding to wavelengths of $\sim$0.9–1.1 Å). The signal to noise ratio attainable from a given dose rises with increasing photon energy, and there are numerous other advantages in using higher energy photons, *e.g.* less absorption (so absorption corrections become largely unnecessary), larger radius of the Ewald sphere (which reduces the size of the blind region and so improves completeness), smaller maximum diffraction angle (so obliquity of incidence on planar detectors is less of an issue), *etc.* An investigation into data collection from HEWL crystals with radiation in the range 6.5–33 keV arose from previous work collecting data on protein crystals in diamond anvil cells for high pressure macromolecular crystallography.[66] That work has been extended with a view to determining the optimum X-ray energy

for crystals of a given size.[67] Using the ratio $(B_f - B_i)/B_i$ or $\Delta B/B$ (where B_f is the Wilson B-factor at the end of data collection and B_i the value at the start), to monitor the effects of radiation damage, the authors showed that $\Delta B/B$ for 6.5 keV X-rays was around an order of magnitude greater than for 33 keV. Monte Carlo simulations were generated which showed that for crystals measuring between ~10–100 μm, the optimum energy is ~24–41 keV. The authors note that the scattering intensity drops off in proportion to $1/E^2$, but this reduction was not considered a problem with modern high brightness synchrotron sources. An additional issue identified is that, although CCD/phosphor detectors operate efficiently at these higher energies, the current generation of pixel array detectors like the Pilatus are based on silicon and are very inefficient at detecting photons with energies remote from ~12 keV.

Over the years, attempts have been made to use chemical free radical scavengers in protein crystals in order to mitigate radiation damage; the results obtained have often been in conflict, and no *panacea* has emerged, either for room temperature or for cryogenic data collection. An in-depth review of the studies published to date has shown that, while part of this disagreement may be due to the different metrics used to quantify the method, it is plain that any effect is small in any case.[68] For example, some studies have followed changes in R_d (based on the difference in intensity of standard reflections during the course of data collection) while others have focussed on ΔB_{rel}, and others on I/I_0; all of these relate to different aspects of the data and are not strictly comparable. R_d is dominated by effects due to non-isomorphism (which is a consequence of radiation damage), ΔB_{rel} is an indication of the increase in overall disorder in the crystal, and I/I_0 is a direct measure of how well the crystal diffracts, and so is more closely related to the quality and resolution of electron density maps. The review ends on a downbeat note; it is apparent that, until some consensus is reached over which metric(s) should be used, it will be difficult to compare studies on even the same scavenger. It was noted, however, that even the best and most reliable scavenger, 1,4-benzoquinone, can only double the crystal lifetime at best.

5 Crystal structure solution

α-Helices occur in roughly 90% of protein molecules; they are characterised by a regular helical structure, with one turn every 3.6 residues, and a pitch of 5.4 Å. This affects the diffraction pattern, and hence the Patterson function, which will have a modulation at multiples of 5.4 Å in projections along the axis of any reasonably long sections of α-helix; by systematically examining the radial profile of the Patterson function in different directions, it should be possible to identify the orientation of this feature in a structure; the Patterson function can be calculated directly from diffracted intensities without phase information but can provide useful information for phasing. An automated search has been developed which makes use of this feature and a number of test cases examined.[69] A Patterson map is calculated from the X-ray data (converted to normalised structure factors, E) using only those reflections with a resolution lower than 3 Å, then projected onto a grid described by spherical polar coordinates, a power spectrum of each projection is calculated, and a peak search is conducted on the power spectra. To improve the discrimination of putative correct solutions,

three stages of selection are employed; stage A contains the list of all candidate solutions found in the initial steps, stage B involves clustering the solutions so that duplicates can be eliminated, and in stage C a figure of merit is calculated for the remaining solutions based on the match of a Patterson projection for a polyalanine α-helix in the same orientation. The overall efficiency of locating the orientations of α-helices was reported to be ~48%. Although no phase information was gained directly, it was noted that the orientation information could be utilised by molecular replacement programs and may help the efficiency of computationally expensive six-dimensional search methods. Knowledge of the expected X-ray diffraction pattern for a helical molecular structure was central to the discovery of the double helical structure of DNA,[7,70,71] so this new method has an important historical precedent.

All quantities measured have uncertainties associated with their values; these variations can be calculated by standard statistical methods and are often presented as variances or as standard deviations. Recent work on examining the properties of these uncertainties has been extended into interpreting their information content, both for small molecules and proteins.[72] If the standard deviations of both the phases and the amplitudes used to calculate electron density are available, and some model is also available (no matter how bad), the variance of the electron density (var $\rho(\boldsymbol{r})=[\sigma_\rho(\boldsymbol{r})]^2$) for each pixel in a given map may be calculated *via* a fast Fourier Transform (FFT), but these values are not usually reported. The variances (and by implication the standard deviations, $\sigma_d(\boldsymbol{r})$) for the individual volume elements in calculated electron density maps also contain useful information; these values are constants for each map and do not vary from pixel to pixel; further, it can be shown that $\sigma_\rho(\boldsymbol{r})$ and $\sigma_d(\boldsymbol{r})$ are anti-correlated, and it can also be shown that the ratio $\rho(\boldsymbol{r})/\sigma_\rho(\boldsymbol{r})$ can be seen as a signal to noise ratio (S/N). Plots of the fraction of pixels in a map with calculated values of S/N have certain features. Firstly, they have a peak at S/N = 0, which corresponds to flat regions of the map (where there are either no atoms or only disordered atoms). Secondly, the plots are asymmetric, and this asymmetry becomes greater with increasing quality of the model. The higher values of S/N correspond with electron density at atomic positions. Thirdly, when the map is bad, S/N rarely exceeds values more than 5, so this can be used as a diagnostic. Fourthly, when heavy atoms are present, S/N may reach large (>100) values; this effect is enhanced with higher resolution data. A cumulative distribution function plot of S/N has a shallower slope as the phase error for the reflections decreases. Hence, an analysis of the variance of the pixel values in an electron density map gives an estimate of confidence in the map.

For the statistician, any analysis of variance immediately raises the question of what is happening to the covariance, and what this can tell us. The covariance gives information on how strongly coupled one feature is to another (in the example of least-squares structure refinement, analysis of the variance–covariance matrix gives the standard deviations of bond lengths and angles). While related, correlation and covariance are not identical; the correlation of two random variables x and y is given by [covariance$(x,y)/\sigma_x\sigma_y$]. An electron density map which has pixels with high covariance indicates that those pixels do not describe independent parameters, and are often associated with unwanted features in the map (*e.g.* incorrect atom placement or other parameters); conversely, maps with

low covariance may be more confidently asserted to be derived from structure factors with small phase errors. A study into the covariances of a one-dimensional electron density map has shown that they are affected by both the Patterson map and uncertainties in the individual measurements.[73] High covariances are found close to large peaks in the map, and this is to be expected, since the electron density should decrease smoothly (if rapidly) away from the maximum for any given atom; a large central value will necessarily be adjacent to significantly large values. Covariance estimates for so-called ''hybrid electron-density maps'', which are some combination of calculated and difference density (see below) show many of the same features as those for calculated electron density maps.

Traditional direct methods have been extremely successful in solving small to medium-sized molecular crystal structures (say, up to $\sim$200 non-hydrogen atoms in favourable cases), while dual-space methods developed since the early 1990's have extended routine structure solution to $\sim$1000 atoms. This success has meant that there has been little incentive to develop the methods involving triplet phase invariants. More recently, the charge flipping algorithm has been successful,[74] as have Patterson deconvolution techniques.[75] The original work on triplet invariants was derived in the absence of knowledge of a structural model, and implied the use of a single X-ray dataset. If some prior structural information is available, or a second dataset exists for some derivative of the ''native'' structure, then this provides further information for obtaining phases. A recent report developed a generalisation of the mathematics, taking into account non-isomorphism between datasets, and showed that, in these circumstances, the joint probability distribution of the triplet phases is shifted relative to that for a single dataset.[76]

A method for solving crystal structures from a set of random phases has been reported which develops the ''Electron Density Modification-Difference Electron Density Modification'' (EDM-DEDM) which came from the same group.[77] The ''Vive la difference'' (VLD) method iterates cycles of EDM-DEDM and drives the structure to correct solutions by including an electron density ''flipping'' cycle of the kind used in charge flipping; in its original incarnation, the method used what the authors term (1−0) and (1−1) syntheses (for the observed and difference electron densities respectively, derived from ''target'' and ''model'' structures); the reader may be more familiar with the terms ''F_o'' and ''$F_o - F_c$'' syntheses (or ''obs maps'' and ''difference maps''). Hybrid syntheses have properties intermediate between the classical (1−0) and (1−1) approaches (*e.g.* true peaks are enhanced, while false peaks are reduced), and these give better discrimination of results. Re-examination of the method in the light of joint probability distribution functions developed earlier allowed extension to *e.g.* (2−1), (1–2) and (1−3) syntheses (*i.e.* $2F_o - F_c$, $F_o - 2F_c$, $F_o - 3F_c$, respectively).[78] VLD was shown to be effective in solving structures ranging in size from small molecules to proteins. In contrast to the more established charge-flipping algorithms, in which higher success has been found when the symmetry is expanded from the true space group to *P*1, the current method gives more reliable results when the correct space group is used in the calculations.

Naturally occurring proteins only crystallise in the 65 chiral space groups, *i.e.* those that do not contain inversion centres, mirror planes or glide planes, since protein chains naturally consist of only the L-enantiomers of each constituent

chiral amino acid. This complicates structure solution because the structure factors can have phases with any value in the range 0–2π. The phases of crystal structures in the remaining 165 achiral space groups can only take one of two values – either 0 or π, so in principle structure solution is more straightforward. For smaller proteins, it is possible to chemically synthesise the other enantiomer (containing only D-amino acids) and co-crystallise the natural and synthetic components into a racemate, and at least one successful structure solution has been performed on such a sample.[79] Phasing may still be difficult, and in a proof of principle experiment, snow flea antifreeze protein was modified to replace a single sulfur with a selenium, and co-crystallised with the non-selenium containing D-amino acid enantiomer to form a quasi-racemate, differing only in a single atom per protein molecule, and for which the reflection phases were approximately 0 or π.[80] Using anomalous differences from a single wavelength dataset (using the SAD method), it was possible to locate the selenium atoms. Two limiting approaches were taken to solve the structure; in the first, the protein phases of 0 or π were taken as ''strong constraints'', with any ambiguity resolved by reference to the phases from the heavy atoms; in the second (more successful) strategy, the phases were derived from the heavy atoms and the ambiguity resolved by determining whether they were closer to 0 or π. The optimum approach might be expected to stem from a choice that balanced these two extremes in a statistically robust fashion, but it was noted that this would be difficult *a priori*, and in the present study could only be done in the light of previous structural knowledge.

It was reported nearly two decades ago that a sphere of electron density corresponding roughly to the molecular envelope of a macromolecule can be placed correctly by ''walking'' the sphere systematically through the asymmetric unit, and that this could give useful phase information, even if only low resolution X-ray data are available.[81] Low resolution information about the spherical envelope can be expanded *via* a series of spherical harmonics to give a better approximation to the molecular shape, using data obtained from small angle X-ray scattering (SAXS); however, it has not generally been possible to expand this information to a molecular structure even if high resolution single crystal X-ray data are available. The technique of oversampling is available if the protein crystal under investigation has a solvent content $>50\%$ and the solvent region is assumed to have uniformly zero electron density (remember that protein crystals contain a significant fraction of water by volume); a method involving dual-space recycling (*i.e.* alternating between real space calculations of electron density and reciprocal space calculations of structure factors) has been shown to be able to extend phase information to ~ 2 Å.[82] This resolution allows chain tracing and the identification of most residues, provided the solvent content is at least 65%. The robustness of the method could be improved by employing density modification by histogram matching. The very low resolution reflections (which are normally not collected as they are obscured by the beamstop) are particularly important in this kind of approach; it was possible to reconstruct approximations of the intensities of these missing reflections by analysis of the data that could be collected, and since it was found that the algorithm was not particularly sensitive to the accuracy of any of the measured structure factor amplitudes, this was found to be a workable approach.

6 Model building and refinement

Automated model building is well-established in both small-molecule and protein crystallography, but is not routine for nucleic acid structures such as DNA and RNA. A method for the location of the phosphate groups in the main chains using a peak-picking approach (developed from one used in small molecule crystallography) followed by real space refinement (RSR) and rejection of false positives (by comparison with known geometry) has been combined with a fragment search for idealised models based on almost complete nucleosides containing either pyrimidine or purine rings bound to ribose rings typical of either A-RNA or B-DNA (the most common forms of each type of nucleic acid); again, this was followed by RSR to optimise the location and orientation.[83]

An alternative semi-automatic approach combines user interaction *via* the graphical model-building program *Coot*,[84] with automated placing of nucleosides using the program plug-in *RCrane*.[85] Provided the initial phases are of sufficient quality, electron density maps will have peaks corresponding to phosphate groups in the main chain backbone. The user picks the first manually, then the plug-in searches within a radius of 10 Å for any large peaks in the electron density maps corresponding to likely candidates for the next phosphate; a choice is given to the user as to which of the plug-in's suggestions to accept. As each phosphate is chosen, the program traces a potential location for the next base and phosphate, at which point the user can accept or reject the choice. Once the chain trace is complete, *RCrane* picks a set of six conformers for each nucleoside in the chain from a library of likely candidates and assigns a score to each. The atomic coordinates of all the alternatives are calculated and a minimisation is performed to obtain the best fit to the observed electron density. At this stage the user can review the program's choices and modify if necessary. *RCrane* was compared with the established and reliable protein auto-building program *ARP/wARP*, which has recently been modified to build nucleic acids as well.[86] It was found that, although *ARP/wARP* was useful in autotracing RNA in protein–nucleic acid complexes, with purely RNA structures the geometries of the nucleotides were grossly distorted. It was suggested that this may be due to the overall low quality of the maps of RNA structures; the methods for fitting protein chains to maps are much better developed, and once the protein has been fitted, the phases are relatively close to being correct, which in turn gives better density for the nucleic acid. The efficacy of *RCrane* was reported as being more sensitive to the quality of the phases than to the resolution of the data, and worked effectively with data extending to only 4 Å.

Molecular replacement often yields a solution from a model which, although it has a high degree of identity to the target, has some significant difference, *e.g.* there may be two domains oriented differently because of a movement in a hinge region; in this scenario, one or other domain may map well to the target, but not both simultaneously. In order to drive the structure solution towards completion, a method has been developed which makes use of local similarities. The process, called ''morphing'' deforms the model iteratively to improve the fit with calculated electron-density.[87] Initially, an electron density map that extends over 11 amino acid residues is calculated, and the optimum translation towards the electron density is applied to the central amino acid in this group. The shifts are smoothed across the whole structure using a linear interpolation based on

the 11 residues, so the central residue is moved most in this stage as well; the new model is refined to improve its geometry and fit to the crystallographic data. It has been found that in a number of test cases, six cycles of morphing were sufficient to yield a sufficiently good model for existing automated building procedures to work. It was noted by the authors that some structures can move sufficiently for subsequent automated building without morphing, but instead using many cycles of conventional refinement, and that this was much slower and less likely to yield an answer. The key to its success compared to existing methods seems to be associated with its ability to apply larger distortions to the model than other protocols which seek to maintain geometry more conservatively.

It is apparent from any crystal structure determination (whether for small or macromolecule single crystals or for powders) that some observations will have more effect on the refinement than others; this influence of an observation is called ''leverage'' and its analysis offers an insight into the importance of both individual data and different classes of data. In particular, a proper analysis of leverage means that the effects of observed and derived observations can be quantified. A study analysing the effect of leverage by a variety of factors has been published; these included structural restraints (which can be considered as derived observations), the effects of incomplete datasets (as might be obtained from data collected in a high pressure study in a diamond anvil cell), and the use of leverage in weighting schemes.[88] The effect on the confidence of the value of the Flack parameter (used for determination of absolute structure) for organic compounds containing no atoms heavier than oxygen was also considered. It was noted that the analysis could be approached from two directions; either by considering the effect of a class of observations on particular structural parameters, or how groups of parameters are affected by individual observations. It was observed that restraints which have little effect on the refinement (*i.e.* leverage ~ 0) serve little purpose, while those that have a large effect may have more effect on refined parameters than the experimental observations. Since different results were obtained for similar compounds using similar restraints, the point is reinforced that care needs to be exercised when applying restraints. The Flack parameter was shown to be more dependent on low resolution (strong) reflections, and high resolution reflections had low leverage; this was explained by the high resolution reflections being weaker and the differences in Bijvoet pairs being harder to measure accurately.

Restraints applied in a refinement should provide extra structural information to make the derived model more physically realistic without a consequent significant increase to the least-squares residuals. A typical restraint can be applied to any parameter that has ''suspicious'' values. Typically, even if one of the atoms involved in a distance restraint is much better defined than the other (say, a heavy atom bound to a disordered ligand), a similar weighting is applied by the refinement programs to each of the atoms, even if the influence of one on the other is significantly greater than *vice versa*. This problem has been addressed by allowing the application of the restraint asymmetrically in the least-squares process in the program *CRYSTALS*.[89] Bonds are taken to be rigid, and the only movement allowed is perpendicular to the direction of the bond. This may be formulated so that the number of restraints that can be applied to an anisotropically refined atom rises to about 3; extending this to include 1,3 interactions

(*i.e.* two atoms that are bonded to a common atom) means that the number of restraints is, on average, just over 6.[90] The implication of this with regard to structures with lower than atomic resolution data is that refinement of anisotropic displacement parameters can be performed when the data-to-parameter ratio is too low for unrestrained refinement. The method was compared with TLS (translation–libration–screw, describing the three modes of anisotropic movement allowed) analysis, a way of reducing the number of parameters by apportioning the movement of atoms in groups between isotropic displacement of the individual atoms and movement of the groups.

The use of restraints is justified where their exclusion leads to over-parameterisation of a refinement or where the model obtained has gross geometrical defects. The former problem is less of an issue where the experimental data extend to atomic resolution or beyond (say, better than 1.2 Å), and in this instance unrestrained refinement may be appropriate. However, geometrical restraints may still be necessary when some parts of the molecule in question have multiple conformations. A method of "trial unrestrained refinement" (TUR) has been suggested for identifying which residues in a protein model may be present in more than one conformation.[91] In the method, a refined structure is subjected to a few cycles of unrestrained refinement and the shifts of the residues from their starting positions analysed; a simple plot of the shifts against residue gives a qualitative indication of the crystallographic order. This was quantified by considering whether a shift of an individual residue is likely to be associated with multiple conformations by two approaches, one involving a simple threshold value for the shift, and the other using a likelihood-based approach, based on the different distribution of shifts for residues with either single or multiple conformations. The results presented show little difference between the two approaches; statistically this implies that the threshold values were chosen sensibly, so either method could be used in predicting which residues were likely to be disordered. The authors found that there appears to be significant under-reporting of multiple conformations in the PDB, even for high resolution structures where identification might be possible. Suggested reasons for this included lack of direct evidence in the electron density maps and the supposition that only a single conformation exists. Interestingly, while the best predictions were made for side chains inside the globular structure of the molecule, and worse predictions occurred for side chains on the surface, much worse results were made for main chain atoms that might be thought to be particularly well-ordered. This was explained by the difficulty of modelling alternative conformations in the main chain.

Most macromolecular refinement programs use restraint targets based on a standard bond length and angle analysis of well-determined small molecule structures contained in the Cambridge Structural Database (CSD).[92] A refinement program *PrimeX* has been described which uses an alternative set of targets based on the parameterisation of van der Waals radii expressed as Lennard-Jones potentials.[93,94] It was suggested that this would be less likely to result in structures which contain severe clashes between atoms, *i.e.* contacts which are less than 0.7× the sum of the atoms' van der Waals radii, while the crystallographic residuals R and R_{free} are not increased significantly. An additional tool in *PrimeX* optimises the structure further by analysing hydrogen bond donor and acceptor locations and minimising the energy of the system by maximising the number of hydrogen bonds,

while avoiding close high-energy non-bonding interactions; the initial solutions identified were minimised using simulated annealing.

The Hamilton *R* test is well-established in crystallography for determining whether additional parameters in a model result in a real improvement, or just give rise to over-fitting. An alternative approach that couches the problem in terms of a Bayesian analysis has been published with respect to the evidence for the presence of minor components in photocrystallography.[95] Compounds that are especially amenable to this kind of experiment are those subject to photo-isomerisation, since these do not require additional reagents for the chemical change. Several different experiments were performed to test for the reliable existence of two different metastable states in the presence of the ground state isomer and the significance evaluated. For example, in the first experiment, a crystal of aquatetraammine(sulfur dioxide)-ruthenium(II) (±)-camphorsulfonate, $[Ru(SO_2)(NH_3)_4(H_2O)](C_{10}H_{15}SO_3)_2$, was analysed before and after irradiation with light; the Bayesian analysis showed that there was little likelihood of the presence of small percentages of the metastable states in the ''dark'' sample, whereas the irradiated sample showed clear evidence. While this was encouraging, the authors took care to note that their analysis was based on a simplified formulation and that this could be improved. They also compared their results with those from the Hamilton *R* test and found that, while comparable for some of their data, small differences were evident that could lead to different conclusions about the experimental results. The reason for this was identified as the two types of test posing different questions; the Bayesian approach compares the likelihood of two different models, whereas the Hamilton test asks whether, given a null hypothesis, there is a substantial chance of observing the same likelihood ratio for the alternative as that actually observed. It should be noted here that Bayesian methods and maximum likelihood approaches are already well established in powder and macromolecular crystallography where the data are mostly sparse compared with small molecule single crystal methods.

The effects of high resolution limit and data completeness on the structures of four small proteins have been investigated with the aim of determining the protonation state of specific amino acid residues.[96] The datasets (cMyBP-C, a cardiac muscle protein 1.3 Å data 88.4% complete, trypsin, 1.2 Å, 97%, subtilisin, 1.26 Å, 100%, and lysozyme 0.65 Å, 98%) were all of sufficiently high resolution and of sufficient completeness that the structures could be refined *via* full matrix least-squares using the program *SHELXL*,[97] and as a result standard uncertainties (often called estimated standard deviations, or e.s.d.s) on the bond lengths and angles could be obtained from the variance–covariance matrices. The protonation state of an atom can be inferred from its bond lengths, provided these can be determined with sufficient accuracy and precision. The authors showed that an increase in resolution of 0.04 Å (from 1.30 to 1.26 Å) and an increase in completeness from 88.4% to 100% taken together were associated with a halving of the e.s.d.s of the bond lengths (comparing cMyBP-C with subtilisin), and a further increase in resolution to 1.20 Å halved the e.s.d.s again (comparing with trypsin), but they were careful to note the difficulty in comparing values from different proteins and different resolutions. However, distinguishing the appropriate protonation state was found to be more reliable as the resolution increased. Perhaps surprisingly, the bond angles in cyclic residues such as

histidine could not be determined with sufficient precision even with 0.65 Å data to allow correct assignment.

7 Summary

A testament to the developers of robust and reliable software for structure solution is that all three major crystallographic databases continue to grow at remarkable rates. As of November 2012, the Cambridge Structural Database (CSD, http://www.ccdc.cam.ac.uk) had ~625 000 entries, the Inorganic Crystal Structure Database (ICSD, http://www.fiz-karlsruhe.de/icsd_content.html) ~140 000 and the Protein Data Bank (PDB, http://www.rcsb.org/pdb/home/home.do), ~87 000 (*cf.* in 2009 the figures were ~500 000 for CSD and 62 500 for PDB in 2009). The free access (both in the sense that it costs nothing and is available without complex licensing) Crystallographic Open Database (http://www.crystallography.net/) currently has ~220 000 entries (*cf.* 88 000 in 2009), so has grown by more individual entries than the CSD.

The bulk of entries in the CSD are not freely available to non-subscribers, but the data from which the entries are drawn are available, in large part, on the worldwide web (WWW). An automated process for aggregating crystallographic data, *CrystalEye*, has been developed and has been used to scan the supplementary material (in CIF format) from various journal publishers that is available on the WWW.[98] The data are written to an XML based repository, and additional information added in Chemical Markup Language (CML). By 2011 *CrystalEye* had processed some 150 000 compound CIFs, containing around 250 000 individual structures.

The rate of increased deposition of structures to the PDB has been discussed recently,[99] in an ''analytically reflective'' note, where the author addressed a number of points that may have affected the growth; unfortunately, he was unable to come to firm conclusions (beyond the observation that the increase in number of total depositions has a natural logarithmic trend, from 1972) but did point out the exponent for the growth equation has barely changed in 40 years; the doubling time has always been in the range ~2.7–3.7 years.

The last report in this series indicated that the statistical approaches used for some time in macromolecular crystallography (*e.g.* Bayesian statistics, maximum likelihood methods, use of ''flat'' regions of electron density) were beginning to be used in powder crystallography; the two fields share the feature of being data-poor (either because of single crystals that only diffract to low resolution, or because of overlap of the diffracted scattering vectors) but information-rich (much is known beforehand about the structures under investigation).[100] There seems to be little evidence for this in the year's literature, although there has been some work published on the use of better statistical methods in small molecule single crystal photochemistry. The major advance of note has been in the introduction of fast pixel array detectors at many synchrotron beamlines. It remains to be seen if the X-ray free electron laser will prove to be a useful new radiation source of general applicability, or if its main use will be in other fields. To a large extent, improvements to both structure solution and refinement have become incremental, so it has become easier and faster to solve and validate structures than a few years ago, but there have been no great leaps forward. Structure solution itself has become more routine in small molecule, protein and

powder crystallography, and it seems that any new major developments will appear in more exotic experiments, such as under extreme conditions or in time-resolved studies.

References

1 M. von Laue, W. Friedrich, P. Knipping, *Sitzungsberichte der Königlich Bayerischen Akademie der Wissenschaften Mathematisch-physikalische Klasse*, 1912, 303–322.
2 M. von Laue, *Sitzungsberichte der Königlich Bayerischen Akademie der Wissenschaften Mathematisch-physikalische Klasse*, 1912, 363–373.
3 W. L. Bragg, *Proc. Cambridge Philos. Soc.*, 1913, **17**, 43–57.
4 R. G. Dickinson and A. L. Raymond, *J. Am. Chem. Soc.*, 1923, **45**, 22–29.
5 J. D. Bernal and D. Crowfoot, *Nature*, 1934, **133**, 794–795.
6 J. C. Kendrew, G. Bodo, H. M. Dintzis, R. G. Parrish, H. Wyckoff and D. C. Phillips, *Nature*, 1958, **181**, 662–666.
7 R. E. Franklin and R. G. Gosling, *Nature*, 1953, **171**, 740–741.
8 A. H. J. Wang, G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. Van der Marel and A. Rich, *Nature*, 1979, **282**, 680–686.
9 Laue centennial special issue, *Acta Crystallogr., Sect. A: Fundam. Crystallogr.*, 2012, **68**, 1–67.
10 Bragg centennial special issue, *Acta Crystallogr., Sect. A: Fundam. Crystallogr.*, 2013, **69**, 1–62.
11 See http://journals.iucr.org/a/issues/2012/01/00/issconts.html and http://journals.iucr.org/a/issues/2013/01/00/issconts.html for online access to these issues.
12 R. A. Judge, E. H. Snell and M. J. van der Woerd, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2005, **61**, 763–771.
13 Y. Yu, X. Wang, D. Oberthür, A. Meyer, M. Perbrandt, L. Duan and Q. Kang, *J. Appl. Crystallogr.*, 2012, **45**, 53–60.
14 E. M. Landau and J. P. Rosenbusch, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 14532–14535.
15 M. Aherne, J. A. Lyons and M. Caffrey, *J. Appl. Crystallogr.*, 2012, **45**, 1330–1333.
16 K. Wako, K. Kimura, Y. Yamamoto, T. Sawaura, M. Shen, M. Tachibana and K. Kojima, *J. Appl. Crystallogr.*, 2012, **45**, 1009–1014.
17 R. G. Sierra, H. Laksmono, J. Kern, R. Tran, J. Hattne, R. Alonso-Mori, B. Lassalle-Kaiser, C. Glöckner, J. Hellmich, D. W. Schafer, N. Echols, R. J. Gildea, R. W. Grosse-Kunstleve, J. Sellberg, T. A. McQueen, A. R. Fry, M. M. Messerschmidt, A. Miahnahri, M. M. Seibert, C. Y. Hampton, D. Starodub, N. D. Loh, D. Sokaras, T.-C. Weng, P. H. Zwart, P. Glatzel, D. Milathianaki, W. E. White, P. D. Adams, G. J. Williams, S. Boutet, A. Zouni, J. Messinger, N. K. Sauter, U. Bergmann, J. Yano, V. K. Yachandrab and M. J. Bogana, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2012, **68**, 1584–1587.
18 L. Lomb, J. Steinbrener, S. Bari, D. Beisel, D. Berndt, C. Kieser, M. Lukat, N. Neef and R. L. Shoeman, *J. Appl. Crystallogr.*, 2012, **45**, 674–678.
19 T. A. White, R. A. Kirian, A. V. Martin, A. Aquila, K. Nass, A. Barty and H. N. Chapman, *J. Appl. Crystallogr.*, 2012, **45**, 331–341.
20 A. J. M. Duisenberg, *J. Appl. Crystallogr.*, 1992, **25**, 92–96.
21 A. G. W. Leslie, H. R. Powell, *Evolving Methods for Macromolecular Crystallography*, 2007, vol. 245, pp. 41–51, ISBN 978-1-4020-6314-5.
22 S. Boutet, L. Lomb, G. J. Williams, T. R. M. Barends, A. Aquila, R. Doak, U. Weierstall, D. P. DePonte, J. Steinbrener, R. L. Shoeman, M. Messerschmidt, A. Barty, T. A. White, S. Kassemeyer, R. A. Kirian, M. M. Seibert, P. A. Montanez, C. Kenney, R. Herbst, P. Hart, J. Pines, G. Haller, S. M. Gruner, H. T. Philipp, M. W. Tate, M. Hromalik, L. J. Koerner, N. van Bakel, J. Morse, W. Ghonsalves, D. Arnlund, M. J. Bogan, C. Caleman, R. Fromme, C. Y. Hampton, M. S. Hunter, L. C. Johansson, G. Katona, C. Kupitz, M. Liang, A. V. Martin, K. Nass, L. Redecke, F. Stellato, N. Timneanu, D. Wang, N. A. Zatsepin, D. Schafer, J. Defever, R. Neutze, P. Fromme, J. C. H. Spence, H. N. Chapman and I. Schlichting, *Science*, 2012, **337**, 362–364.
23 J. Hajdu, *Curr. Opin. Struct. Biol.*, 2000, **10**, 569–573.
24 A. Tokuhisa, J. Taka, H. Kono and N. Go, *Acta Crystallogr., Sect. A: Fundam. Crystallogr.*, 2012, **68**, 366–381.
25 U. Mueller, N. Darowski, M. R. Fuchs, R. Förster, M. Hellmig, K. S. Paithankar, S. Pühringer, M. Steffien, G. Zocher and M. S. Weiss, *J. Synchrotron Radiat.*, 2012, **19**, 442–449.
26 P. Trueb, B. A. Sobott, R. Schnyder, T. Loeliger, M. Schneebeli, M. Kobas, R. P. Rassool, D. J. Peake and C. Broennimann, *J. Synchrotron Radiat.*, 2012, **19**, 347–351.

27 Ch. Broennimann, E. F. Eikenberry, B. Henrich, R. Horisberger, G. Huelsen, E. Pohl, B. Schmitt, C. Schulze-Briese, M. Suzuki, T. Tomizaki, H. Toyokawa and A. Wagner, *J. Synchrotron Radiat.*, 2006, **13**, 120–130.
28 C. Le Bourlot, P. Landois, S. Djaziri, P.-O. Renault, E. Le Bourhis, P. Goudeau, M. Pinault, M. Mayne-L'Hermite, B. Bacroix, D. Faurie, O. Castelnau, P. Launois and S. Rouzière, *J. Synchrotron Radiat.*, 2012, **19**, 38–47.
29 J. M. Olton, C. Nielsen and K. A. Frankel, *J. Synchrotron Radiat.*, 2012, **19**, 1006–1011.
30 J. M. Holton, *Acta Crystallogr., Sect. A: Fundam. Crystallogr.*, 2008, **64**, C77.
31 A. G. W. Leslie, H. R. Powell, G. Winter, O. Svensson, D. Spruce, S. McSweeney, D. Love, S. Kinder, E. Duke and C. Nave, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2002, **58**, 1924–1928.
32 M.-F. Incardona, G. P. Bourenkov, K. Levik, R. A. Pieritz, A. N. Popov and O. Svensson, *J. Synchrotron Radiat.*, 2009, **16**, 872–879.
33 S. Brockhauser, O. Svensson, M. W. Bowler, M. Nanao, E. Gordon, R. M. F. Leal, A. Popov, M. Gerring, A. A. McCarthy and A. Gotz, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2012, **68**, 975–984.
34 M. Mueller, M. Wang and C. Schulze-Briese, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2012, **68**, 42–56.
35 W. Kabsch, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 125–132.
36 J. A. Coome, A. E. Goeta, J. A. K. Howard and M. R. Probert, *J. Appl. Crystallogr.*, 2012, **45**, 292–298.
37 e.g. T. Skarzynski, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2009, **65**, s159.
38 D. Axford, R. L. Owen, J. Aishima, J. Foadi, A. W. Morgan, J. I. Robinson, J. E. Nettleship, R. J. Owens, I. Moraes, E. E. Fry, J. M. Grimes, K. Harlos, A. Kotecha, J. Ren, G. Sutton, T. S. Walter, D. I. Stuart and G. Evans, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2012, **68**, 592–600.
39 D. Hargreaves, *J. Appl. Crystallogr.*, 2012, **45**, 138–140.
40 M. Sugahara, *J. Appl. Crystallogr.*, 2012, **45**, 362–366.
41 M. Makino, I. Wada, N. Mizuno, K. Hirata, N. Shimizu, T. Hikima, M. Yamamoto and T. Kumusaka, *J. Appl. Crystallogr.*, 2012, **45**, 785–788.
42 M. Kurz, B. Blattmann, A. Kaech, C. Briand, P. Reardon, U. Ziegler and M. G. Gruetter, *J. Appl. Crystallogr.*, 2012, **45**, 999–1008.
43 T. A. Pemberton, B. R. Still, E. M. Christensen, H. Singh, D. Srivastava and J. J. Tanner, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2012, **68**, 1010–1018.
44 D. Waterman and G. Evans, *J. Appl. Crystallogr.*, 2010, **43**, 1356–1371.
45 D. Liebschner, M. Dauter, G. Rosenbaum and Z. Dauter, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2012, **68**, 1430–1436.
46 K. Diederichs, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 733–740.
47 E. Pozharski, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2012, **68**, 1077–1087.
48 Z. Otwinowski and W. Minor, *Methods Enzymol.*, 1997, **276**, 307–326.
49 P. Evans, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2006, **62**, 72–82.
50 F. H. C. Crick and B. S. Magdoff, *Acta Crystallogr.*, 1956, **9**, 901–908.
51 P. A. Karplus and K. Diederichs, *Science*, 2012, **336**, 1030–1033.
52 K. Diederichs, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2013, **69**, DOI: 10.1107/S0907444913001121.
53 K. Diederichs, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2009, **65**, 535–542.
54 J. Henn and K. Meindl, *Acta Crystallogr., Sect. A: Fundam. Crystallogr.*, 2010, **66**, 676–684.
55 M. R. V. Jørgensen, H. Svendsen, M. S. Schmøkel, J. Overgaard and B. B. Iversen, *Acta Crystallogr., Sect. A: Fundam. Crystallogr.*, 2012, **68**, 301–303.
56 J. Henn and K. Meindl, *Acta Crystallogr., Sect. A: Fundam. Crystallogr.*, 2012, **68**, 304.
57 M. Krug, M. S. Weiss, U. Heinemann and U. Mueller, *J. Appl. Crystallogr.*, 2012, **45**, 568–572.
58 U. Schmueli, *J. Appl. Crystallogr.*, 2012, **45**, 389–392.
59 V. Y. Lunin, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2012, **68**, 732–734.
60 R. B. G. Ravelli, H.-K. S. Leiros, B. Pan, M. Caffrey and S. McSweeney, *Structure*, 2003, **11**, 217–224.
61 D. de Sanctis and M. H. Nanao, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2012, **68**, 1152–1162.
62 R. L. Owen, E. Rudiño-Piñera and E. F. Garman, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 4912–4917.
63 M. Warkentin, R. Badeau, J. B. Hopkins, A. M. Mulichak, L. J. Keefe and R. E. Thorne, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2012, **68**, 124–133.
64 M. Warkentin, R. Badeau, J. B. Hopkins and R. E. Thorne, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2012, **68**, 1108–1117.

65 R. L. Owen, D. Axford, J. E. Nettleship, R. J. Owens, J. I. Robinson, A. W. Morgan, A. S. Doré, G. Lebon, C. G. Tate, E. E. Fry, J. Ren, D. I. Stuart and G. Evans, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2012, **68**, 810–818.
66 R. Fourme, E. Girard, A.-C. Dhaussy, K. Medjoubi, T. Prangé, I. Ascone, M. Mezouar and R. Kahn, *J. Synchrotron Radiat.*, 2011, **18**, 13–36.
67 R. Fourme, V. Honkimäki, E. Girard, K. Medjoubi, A.-C. Dhaussyd and R. Kahn, *J. Appl. Crystallogr.*, 2012, **45**, 652–661.
68 E. G. Allan, M. C. Kander, I. Carmichael and E. F. Garman, *J. Synchrotron Radiat.*, 2013, **20**, 23–36.
69 R. Caliandro, D. Dibenedetto, G. L. Cascarano, A. Mazzonea and G. Nico, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2012, **68**, 1–12.
70 J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 737–738.
71 M. H. F. Wilkins, A. R. Stokes and H. R. Wilson, *Nature*, 1953, **171**, 738–740.
72 C. Giacovazzo and A. Mazzone, *Acta Crystallogr., Sect. A: Fundam. Crystallogr.*, 2012, **68**, 464–469.
73 A. Altomare, C. Cuocci, C. Giacovazzo, A. Moliterni and R. Rizzi, *Acta Crystallogr., Sect. A: Fundam. Crystallogr.*, 2012, **68**, 244–255.
74 G. Oszlányi and A. Sütő, *Acta Crystallogr., Sect. A: Fundam. Crystallogr.*, 2005, **61**, 147–152.
75 J. W. Richardson and R. A. Jacobson, in *Patterson and Pattersons*, ed. J. P. Glusker, B. K. Patterson and M. Rossi, Oxford University Press, 1987, pp. 310–317.
76 M. C. Burla, B. Carrozzini, G. L. Cascarano, G. Comunale, C. Giacovazzo, A. Mazzone and G. Polidori, *Acta Crystallogr., Sect. A: Fundam. Crystallogr.*, 2012, **68**, 513–520.
77 R. Caliandro, B. Carrozzini, G. L. Cascarano, L. De Caro, C. Giacovazzo and D. Siliqi, *Acta Crystallogr., Sect. A: Fundam. Crystallogr.*, 2008, **64**, 519–528.
78 M. C. Burla, B. Carrozzini, G. L. Cascarano, C. Giacovazzo and G. Polidori, *J. Appl. Crystallogr.*, 2012, **45**, 1287–1294.
79 J. R. Banigan, K. Mandal, M. R. Sawaya, V. Thammavongsa, A. P. Hendrickx, O. Schneewind, T. O. Yeates and S. B. H. Kent, *Protein Sci.*, 2010, **19**, 1840–1849.
80 M. R. Sawaya, B. L. Pentelute, S. B. H. Kent and T. O. Yeates, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2012, **68**, 62–68.
81 See, for example G. W. Harris, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 1995, **51**, 695–702.
82 Z. C. Liu, R. Xu and Y.-H. Dong, *Acta Crystallogr., Sect. A: Fundam. Crystallogr.*, 2012, **68**, 256–265.
83 F. Pavelčík, *J. Appl. Crystallogr.*, 2012, **45**, 309–315.
84 P. Emsley, B. Lohkamp, W. G. Scott and K. Cowtan, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2010, **66**, 486–501.
85 K. S. Keating and A. M. Pyle, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2012, **68**, 985–995.
86 J. Hattne and V. S. Lamzin, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2008, **64**, 834–842.
87 T. C. Terwilliger, R. J. Read, P. D. Adams, A. T. Brunger, P. V. Afonine, R. W. Grosse-Kunstleve and L.-W. Hung, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2012, **68**, 861–870.
88 S. Parsons, T. Wagner, O. Presly, P. A. Wood and R. I. Cooper, *J. Appl. Crystallogr.*, 2012, **45**, 417–429.
89 R. I. Cooper, A. Thorn and D. J. Watkin, *J. Appl. Crystallogr.*, 2012, **45**, 1057–1060.
90 A. Thorn, B. Dittrich and G. M. Sheldrick, *Acta Crystallogr., Sect. A: Fundam. Crystallogr.*, 2012, **68**, 448–451.
91 O. B. Sobolev and V. Y. Lunin, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2012, **68**, 1118–1127.
92 R. A. Engh and R. Huber, *Acta Crystallogr., Sect. A: Fundam. Crystallogr.*, 1991, **47**, 392–400.
93 R. S. Rowland and R. Taylor, *J. Phys. Chem.*, 1996, **100**, 7384–7391.
94 J. A. Bell, K. L. Ho and R. Farid, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2012, **68**, 935–952.
95 A. E. Philips and J. M. Cole, *Acta Crystallogr., Sect. A: Fundam. Crystallogr.*, 2012, **68**, 324–330.
96 S. J. Fisher, M. P. Blakeley, M. Cianci, S. McSweeney and J. R. Helliwell, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2012, **68**, 800–809.
97 G. M. Sheldrick, *Acta Crystallogr., Sect. A: Fundam. Crystallogr.*, 2008, **64**, 112–122.
98 N. Day, J. Downing, S. Adams, N. W. England and P. Murray-Rust, *J. Appl. Crystallogr.*, 2012, **45**, 316–323.
99 C. Abad-Zapatero, *Acta Crystallogr., Sect. D: Biol. Crystallogr.*, 2012, **68**, 613–617.
100 H. R. Powell, *Annu. Rep. Prog. Chem., Sect. C*, 2010, **106**, 192–210.