

# Current state and prospects of macromolecular crystallography

**Zbigniew Dauter**

Synchrotron Radiation Research Section, MCL,  
National Cancer Institute, Argonne National  
Laboratory, Building 202, 9700 South Cass  
Avenue, Argonne, IL 60439, USA

Correspondence e-mail: [dauter@anl.gov](mailto:dauter@anl.gov)

The current situation and possible future development of macromolecular crystallography are reviewed. The rapid progress and maturation of this field in recent years have to a large extent been made possible by the inspiration and generous support of several active structural genomics initiatives. Two tendencies can be currently observed: one which treats protein crystallography as a highly automatic tool for investigating various biological problems without the need to engage in the intricacies of the technique and a second approach where this method is applied to crystals of difficult, large and complex biological systems, requiring a deeper knowledge of various aspects of crystallography. In the near future it is expected that these two trends will coexist, developing in a parallel fashion.

Received 6 April 2005

Accepted 20 October 2005

## 1. Introduction

Macromolecular crystallography (often traditionally, but less correctly, named protein crystallography, here abbreviated to PX) has undergone tremendous progress in the last few years. Throughout almost 50 years of macromolecular crystallography, development of the crystallographic methodology has progressed mostly in individual academic laboratories, although sometimes it has advanced through the joint efforts of several developers such as the CCP4 initiative in the UK (Collaborative Computational Project, Number 4, 1994). In the first decades of its development, the number of laboratories and researchers practicing PX was limited: every researcher knew everyone else and advances and new programs were usually shared among the whole research community worldwide. Macromolecular crystallography was truly a separate branch of basic science and its practitioners had no inclination towards the patenting of methodology or the commercialization of computer software. Such unselfish attitudes still exist in some laboratories, although they have fallen more and more out of fashion.

The situation changed with the maturation of protein crystallography and the growth of the number of PX laboratories and active researchers, including those in industrial companies. This has been the inevitable result of the steady progress in this field. Whereas in earlier days practicing X-ray diffraction studies of macromolecular crystals required intricate, arcane knowledge of biochemistry, physics, mathematics, engineering and computing, currently protein crystal structures can often be elucidated by the uninitiated layman using available easy-to-use hardware and software systems. This is also a sign of maturity: the method has advanced to the stage

where it often can be treated as one of the available analytical tools for investigating macromolecular structure. This development parallels that of small-molecule crystallography from the heroic first years of the Braggs to the situation where it may be easier and faster to automatically solve the crystal structure of an unknown substance rather than to analyze it by some other method.

The change in stance of PX during the more than 50 years since the time of the giant pioneers such as Perutz, Kendrew, Phillips and Blow can be viewed either as a period of enormous progress 'at a cost' of declining standards (Matthews, 2003) or as a time of maturation of crystallographic methodology and techniques and the deeper involvement of PX in the fulfillment of the needs of human society. Whereas accounts of the early achievements of PX can be read as exciting detective stories (Judson, 1979; Blake *et al.*, 2001; Matthews, 2003; Rossmann, 2003; Arndt, 2003), some (certainly not all) of contemporary PX activity follows the 'publish or perish' rule. The weighting scheme between these two attitudes is not ultimately defined (correct weighting is a recurring problem in PX, anyway).

The progress of macromolecular crystallography in the past 70 years since the recording of the first diffraction pattern from a crystal of pepsin (Bernal & Crowfoot, 1934) has advanced in steps. Therefore, the only true revolution was achieved by Max Perutz when he solved the crystal structure of hemoglobin in the 1950s. At that time, solution of an acentric small structure with less than 100 atoms was very difficult; horse hemoglobin has 2291 non-H atoms in the asymmetric unit. It required a very dedicated and patient researcher (Perutz) and a very involved laboratory chief (Sir W. L. Bragg) to spend nearly 20 years pursuing such a marginally promising project. Today, in a time of strict planning of research goals and tight control of expenditure, such dedication would be unthinkable. We should be immensely grateful to the early believers in the future of protein crystallography for opening the door to our current activities.

The elucidation of the crystal structures of myoglobin (Kendrew *et al.*, 1960) and hemoglobin (Perutz *et al.*, 1960) were certainly a revolution in structural science, rewarded by two Nobel prizes (Kendrew, 1963; Perutz, 1963). Since then, the progress of PX has followed an evolutionary rather than a revolutionary path. Many important advances have been achieved on the way, but most of the methods used by Perutz are still valid today and still useful, albeit often in modified versions.

In the light of this, it is difficult to extrapolate the progress of macromolecular crystallography too boldly into the future. Obviously, one cannot exclude the possibility of a major breakthrough in the methodology, 'There are more things in heaven and earth... than are dreamt of in your philosophy' (Shakespeare, 1603); after all, if such a development were imaginable, it would already be implemented. Too bold statements about the future of science are very risky. For example, the great physicist Lord Kelvin warned in 1899 that 'X-rays will prove to be a hoax' ([http://zapatopi.net/kelvin/](http://zapatopi.net/kelvin/quotes.html)

[quotes.html](http://zapatopi.net/kelvin/quotes.html)). As it turned out, our scientific careers are based on Lord Kelvin being wrong.

## 2. Progress in methodology

It is easiest to realise how much the methods used in macromolecular crystallography have advanced by comparing the several years necessary to solve the first crystal structures of proteins in the 1960s and 1970s with the mere few days sufficient to deposit the refined model of the contemporary structural genomics target after its crystals have been obtained. Although the whole process follows the same stages of crystallization, data collection, phasing, model building and refinement, validation and presentation of the results, today these steps can be performed dramatically more easily and more quickly.

Apart from advances in the strictly crystallographic methodology, very substantial credit has to be given to the enormous progress in computing technology. In fact, the progress in crystallography (both small-molecular and macromolecular) to a large extent parallels the increase in the speed and capacity of computers in the last decades. Some of the algorithms used today would not have been at all possible, even a few years ago. For example, a single magnetic tape used in the 1980s could store 18 diffraction images scanned from photographic films with 100  $\mu\text{m}$  pixels, which corresponded to six exposures with three films in each pack, so that diffraction data had to be processed sequentially in batches of six packs. One image from a contemporary  $3 \times 3$  mosaic CCD detector (in binned mode) occupies about 18 Mb of disk space and it is no problem to integrate millions of individual reflections from several hundred exposures recorded on such a detector in one job. Today, all calculations necessary for crystal structure solution can be performed on a laptop computer.

The complete solution of the three-dimensional crystal structure of a macromolecule still consists of several distinct stages. They are more interlinked and automated than in the past, but all of them are necessary, from sample preparation through crystallization, diffraction data collection, phasing, model building and refinement and validation to the deposition and presentation of the results obtained. These steps will be addressed in the following sections, except for the preparation of the initial biological material, which involves biochemical and genetic techniques and is beyond the scope of the present text. However, it should be very strongly stressed that without the enormous advances in 'wet' molecular biology, the achievements of protein crystallography would be severely limited.

Structural genomics initiatives, relying on the rapid high-throughput elucidation of many structures of proteins from the same genome or class, have played a very important role in the methodological progress of macromolecular crystallography in recent years. Several centers engaged in structural genomics projects in the USA (Terwilliger, 2000b), Japan (Yokohama *et al.*, 2000) and Europe (Heinemann, 2000) actively promote and generously support laboratories and researchers focusing on the development of methods and this

support is in a large part responsible for the recent boom in crystallographic methodology.

### 2.1. Preparation of crystals

Protein crystals were first obtained from bulk solution, an approach known as the batch method. These proteins were available in relatively large amounts from natural sources, *e.g.* urease (Sumner, 1926*a,b*), pepsin (Northrop, 1929, 1930), trypsin (Northrop, 1932), catalase (Sumner & Dounce, 1937*a,b*), TMV nucleoprotein (Stanley, 1935) and lysozyme (Abraham & Robinson, 1937; Alderton *et al.*, 1945). Crystallization of these proteins proved that they are ordinary, albeit large, chemical molecules having well defined three-dimensional structures and for this achievement Sumner, Northrop and Stanley were honored with a Nobel Prize in chemistry in 1946. Since then, many other ways to crystallize proteins (and oligonucleotides) have been proposed (reviewed in McPherson, 1982; Ducruix & Giegé, 1992; Bergfors, 1999), among which the most popular are the vapor-diffusion methods in the hanging-drop (Davies & Segal, 1971) or sitting-drop variants. Other techniques are based on the counterdiffusion (García-Ruiz, 2003) of the protein and precipitant solutions through membranes (Zeppetzauer *et al.*, 1968), in gel (García-Ruiz & Moreno, 1994) or in capillaries (Ng *et al.*, 2003). Very useful for automatic systems is the use of microbatch crystallization (Chayen *et al.*, 1990), also under a layer of inert oil (Chayen *et al.*, 1992).

Macromolecules always require some precipitants to invoke nucleation and crystal growth. The most popular precipitants are inorganic salts (ammonium sulfate, sodium chloride *etc.*) and organic polyols (polyethylene glycol of various sizes, methylpentanediol and similar additives). A multitude of other parameters can be varied, *e.g.* the buffer type and pH, temperature, concentration of protein and other components *etc.* Some proteins require small amounts of special additives such as dioxane, phenol, 2-propanol or various cofactors or effectors to produce good-quality crystals. Various sets of crystallization screening conditions selected by sparse-matrix sampling (Jancarik & Kim, 1991) have been proposed and many ready-made solutions are available commercially (Hampton Research, <http://www.hamptonresearch.com/>; Emerald Biostructures, <http://www.emeraldbiostructures.com/>; Jena Bioscience, <http://www.jenabioscience.com/>; Molecular Dimensions, <http://www.moleculardimensions.com/>).

A recent tendency is to perform the initial crystallization trials in a multitude of conditions with small amounts of biological material, using droplets of very small volume, of the order of tens of nanolitres. Preparation of thousands of small droplets of varying but strictly prescribed composition can be very effectively performed by robots. Indeed, crystallization was one of the first steps of X-ray structure analysis where automation was successfully introduced (Chayen *et al.*, 1990; Weselak *et al.*, 2003) and currently many academic and industrial laboratories are equipped with various crystallization robots. The generally available protein-crystallization service available at Buffalo (<http://www.hwi.buffalo.edu/>

[ProductsServices/highthroughput/HighThrough.htm](http://ProductsServices/highthroughput/HighThrough.htm)) uses robotics to screen for optimum crystallization conditions.

Crystallization robots are often linked to devices that automatically inspect all crystallization setups through an optical microscope and are able to classify the individual drops for the presence of various features, such as amorphous precipitate, microcrystals, sizable crystals *etc.* Several efforts are currently in progress towards the reliable automatic identification of crystals in the drops (and in the mounting loops; see below). The final judgment of the preclassified crystals is usually still performed by the human eye because it is currently not possible to automate the process of fishing out the selected crystals from the drop. This may change in the not too distant future. One proposed way of circumventing this problem is based on growing crystals in special 'matrices' and inserting all setups with drops into the X-ray beam for diffraction experiments (Watanabe *et al.*, 2002; Jacquamet *et al.*, 2004).

For decades, crystals were mounted for diffraction experiments in sealed glass (or quartz) capillaries in equilibrium with the drop of mother liquor, following the first observation of Bernal & Crowfoot (1934) that protein crystals disintegrate and lose diffraction properties after drying out. Such a method of crystal mounting obviously needed to be performed by hand. A breakthrough in protein crystal handling took place after introducing crystal-cryocooling techniques. Using this method, now commonly referred to as cryocrystallography, the mother liquor within the solvent channels of the crystal and surrounding it, with the possible addition of some cryoprotectants, can be solidified in the state of amorphous glass and the sample preserves its crystalline order and diffraction properties (Hope, 1988). The crystal is then scooped from the liquid with a small fiber loop and quickly plunged in the liquid nitrogen or placed in the stream of the cold gas, usually nitrogen at temperatures about 100 K (Teng, 1990). This method of crystal handling exposes them to less mechanical stress and diminishes the amount of radiation damage induced by exposure to X-rays (Garman & Schneider, 1997).

Various modifications of crystal-handling protocols have recently been proposed. The humidity of the atmosphere surrounding crystals can be precisely controlled before cryocooling (Kiefersauer *et al.*, 2000); partial drying out of crystals has been observed to enhance their diffraction properties (Abergel, 2004). In addition, crystal mounting in loops with removal of the surrounding liquid may decrease the absorption effects, especially when a long X-ray wavelength is used (Kitago *et al.*, 2005).

Membrane proteins require the presence of detergents for crystallization (Michel & Oesterhelt, 1980; Garavito & Rosenbusch, 1980); otherwise, they do not dissolve in aqueous solutions and often aggregate, forming micelles. The ingenious way of dealing with membrane proteins is crystallization '*in cubo*', *i.e.* in the scaffold of the cubic lipid phase (Landau & Rosenbusch, 1996; Chiu *et al.*, 2000).

Crystals grown in microgravity are often more perfect than those grown on earth (Vergara *et al.*, 2003). Currently, the high cost and complicated organization of space missions limit the

practicality of this approach, but it may be expected that in the future more crystals will be grown in space.

### 2.2. Diffraction data collection

Measurement of the intensities of diffracted reflections is the last truly experimental step of crystal structure analysis, in the sense that all subsequent stages involve only computational efforts and can be repeated and varied many times once a high-quality data set is available. Crucial technological advances have transformed the diffraction data-acquisition process over the years of the evolution of PX from the most tedious and lengthy stages to one of the easiest and most automated steps of the crystal structure elucidation process. These advances relate to the available X-ray sources, detectors and computer software.

**2.2.1. X-ray sources.** The only X-ray sources available in the early days of protein crystallography were sealed tubes. They are still used in small-molecule crystallography, where the X-ray flux delivered by these devices is usually sufficient. Unfortunately, for weakly diffracting crystals of macromolecules, their output was rather low and they were superseded by rotating-anode machines that deliver more intense X-ray beams (Arndt, 2003). However, Uli Arndt, a hero of the crystallographic X-ray source technology, has recently developed a microfocus sealed tube that is as potent as rotating-anode sources (Arndt, Long *et al.*, 1998).

Currently, rotating-anode machines are the typical laboratory sources of X-rays. The first rotating anodes were somewhat 'temperamental' and required careful maintenance. Contemporary machines are much more stable and powerful. Whereas in small structure crystallography the most popular is the molybdenum target producing X-rays of 0.71 Å wavelength, protein crystals usually do not diffract to subatomic resolution and do not require so short a wavelength. Therefore, copper anodes, providing more flux at 1.54 Å wavelength, are standard in PX. The recently introduced chromium-based ( $\lambda = 2.23$  Å) rotating anode (Yang *et al.*, 2003) appropriately answers the current interest in phasing based on the anomalous scattering signal of sulfur and other relatively light elements.

A substantial part of the responsibility for the increase in the usable X-ray flux of the laboratory X-ray sources arises from improved properties of the optical elements. Precise double-reflecting mirrors and multilayered devices are able to deliver a highly focused and intense X-ray beam at the sample that is comparable to that of second-generation bending-magnet synchrotron beamlines (Arndt, Duncumb *et al.*, 1998).

Of course, the sources of the highest intensity X-rays are synchrotron beamlines. Since the pioneering work of Rosenbaum and Holmes (Rosenbaum *et al.*, 1971; Holmes & Rosenbaum, 1998), these sources have evolved from being exotic and hardly available to the status of the standard and most popular places for collecting diffraction data from macromolecular crystals. Gerd Rosenbaum has very actively contributed to the construction of many of the currently available synchrotron beamlines (Abad-Zapatero, 2004). The

first experiments with the use of synchrotron radiation for recording diffraction images from protein crystals were performed at Stanford (Phillips *et al.*, 1976). In 2004, about 76% of all submissions to the Protein Data Bank were based on data measured at synchrotron beamlines (<http://asdp.bnl.gov/asda/Libraries/>; Jiang & Sweet, 2004). Today, home laboratory sources are often only used for preliminary characterization of crystals before the trip to a synchrotron.

At present, more than 70 active synchrotron beamlines serve the needs of macromolecular crystallographers worldwide (Jiang & Sweet, 2004). Several new synchrotrons are being built and it seems that available synchrotron beam time supply may soon match demand. The availability of synchrotron beams worldwide will be further increased in practice by the expansion of 'FedEx' data-collection facilities (pioneered at NSLS; <http://www.px.nsls.bnl.gov/fedex.html>), where diffraction data are measured by the synchrotron staff scientists on crystals sent from remote locations. Efforts are under way to allow remote users to control the process of data collection through ethernet-based software systems, *e.g.* at Stanford (Sharp, 2004), or to completely automate the collection and reduction of diffraction data (for example, the DNA project in the UK; <http://www.dna.ac.uk/index.html>).

The technology of synchrotron beam facilities is constantly improving. Whereas fixed-wavelength bending-magnet lines were initially the most popular, at present the majority of newly built beamlines are tunable-wavelength lines, often equipped with insertion devices (undulators or wigglers). Third-generation storage rings deliver radiation of high brightness with a highly parallel beam from small-size sources. Such facilities permit the researcher to collect a comprehensive data set in a few minutes and to solve the structure in hours (Walsh *et al.*, 1999). In the future, even more powerful fourth-generation X-ray sources, based on the concept of the free-electron laser (FEL), will be able to deliver beams of coherent X-rays of enormous intensity, several orders of magnitude higher than that available today. The use of such strong radiation sources will present additional challenges to the ways that diffraction data are collected; currently, it is not quite clear whether the sample will first evaporate or produce a measurable diffraction pattern. X-ray FEL facilities are being built at SLAC, Stanford, CA, USA (<http://www-ssrl.slac.stanford.edu/lcls/>) and DESY, Hamburg, Germany (<http://www-hasyllab.desy.de/facility/fel/>) and their future use will open up new horizons for structural biology with the possibility of recording data from very small crystals or even from non-crystalline samples of intra-cell organelles (Miao *et al.*, 1998, 2001; Huldt *et al.*, 2003).

On the opposite scale of magnitude is the concept of a 'desktop' synchrotron that is currently being constructed (<http://www.lynceantech.com/>). It will deliver an X-ray beam comparable with that of contemporary synchrotron facilities and may dramatically change data-collection practices in laboratories in universities and industrial companies.

**2.2.2. Detectors.** At first, protein diffraction data were recorded on photographic films and precession cameras (Buerger, 1964). This was a very tedious procedure and apart

from recording lengthy exposures of diffraction patterns, it required the crystals to be precisely aligned and subsequently involved chemical processing of the exposed films and scanning them on optical densitometers. The introduction of the screenless rotation method (Arndt & Wonacott, 1977) sped up and simplified the process of recording data, but the necessity for film processing still remained. Use of automatic three- and four-circle diffractometers (Arndt & Phillips, 1961; Arndt & Willis, 1966), which revolutionized small-molecule crystallography, was not very practical for collecting data from macromolecular crystals, since the reflection intensities were recorded sequentially, one by one, which was prohibitive because of the incurred radiation damage and data-collection times.

The introduction of automatic computer-controlled two-dimensional detectors, multiwire proportional counters, TV cameras and eventually imaging plates and charge-coupled devices (CCDs) allowed diffraction images to be directly stored on computer disks for further data reduction. These technological advances have not only made the process of data collection more automatic and much faster (in conjunction with the progress in X-ray sources), but have in general increased the level of accuracy of measured diffraction intensities. The currently used CCD detectors can have large front windows, capturing wide diffraction angles, and are characterized by fast read-out times, keeping pace with the short subsecond exposures at bright synchrotron beamlines. In the near future, the prospect of even larger and faster solid-state pixel detectors is on the horizon. The very fast transfer detectors may open up the interesting possibility of recording diffraction images from constantly rotating crystals with exposures controlled by the precisely synchronized shutter (Brönnimann *et al.*, 2004). The availability of very fast detectors may revive the popularity of the fine-slicing method of data collection, leading to more accurate evaluation of reflection profiles and their intensities.

The early precession cameras required crystals to be set precisely along one of the crystal axes. Single-axis rotation cameras do not have this requirement, but also do not provide an easy way of precisely orienting a crystal with respect to the X-ray beam and rotation-axis directions, apart from hand adjustment of the goniometer arcs. Renewed interest in the use of  $\kappa$ -goniostats, and motorized crystal translation stages, make sample orientation and centering much easier and faster.

Contemporary CCD detectors are characterized by high spatial resolution, low intrinsic noise and short read-out time. However, like all two-dimensional detectors, they need to be properly calibrated to faithfully represent the geometry and intensities in all pixels of the recorded diffraction pattern. The accuracy of the estimation of strong intensities can be better than 1%. Since the dynamic range of CCD detectors is limited, attention should be directed to adequate estimation of the most intense reflections, when some detector pixels may be overloaded with the long exposures required to measure weak high-resolution reflections. At present, such a situation requires a separate short-exposure pass, but it would be more

appropriate to use a sort of automatic electronic attenuation or rescaling of such pixels.

**2.2.3. Data-processing software.** The extraction of usable estimations of reflection intensities from the recorded diffraction patterns involves their initial interpretation and indexing, integration of the individual reflection profiles, application of all necessary corrections and subsequent merging and scaling of all measured intensities.

Interpretation and indexing of the early precession photographs was easy, since the recorded patterns represented the undistorted single layers of reflections in the reciprocal lattice. In contrast, in the screenless rotation method, reflections resulting from the cross-section of the reciprocal lattice by the curved surface of the Ewald sphere are represented on the usually flat detector in a distorted way and indexing of such patterns is not trivial. Two basic methods of automatic indexing are in use: the reciprocal-space approach, based on the analysis of the basic set of the inter-peak vectors (Kabsch, 1993), and the currently more popular direct-space approach, based on the interpretation of the Fourier transform of the pattern (Rossmann & van Beek, 1999). This procedure provides the information about the crystal cell and its disposition with respect to the X-ray beam, detector and goniostat axes.

The evaluation of reflection intensities involves the integration of the total intensity within all pixels of the individual reflection profiles. For this, the positions of all reflections potentially present in a given diffraction image have to be predicted from the initial interpretation. The integration may proceed by simple summation of the content of individual pixels within the reflection profile or by a more sophisticated method of fitting the standard profile built from a subset of well measured reflections. At the same time, the uncertainties of the measured intensities are estimated from the analysis of the surrounding background pixels.

The next step is the merging and averaging of the intensities of the multiple measurements of the same or symmetrically equivalent reflections coupled with the simultaneous evaluation of the relative scales of individual diffraction images. Instead of using the simple individual image scales or scales varying smoothly with the spindle-axis rotation, it is possible to use a more elaborate model with scales expressed as a combination of spherical harmonic functions, which better represents the directionality of the sample absorption dependent on the crystal shape (Blessing, 1995).

The whole procedure of data processing is performed by one of the available highly sophisticated software systems, each of which is constantly updated and improved. The most widely used are *MOSFLM* (Leslie, 1992), *HKL2000* (Otwinowski & Minor, 1997), *XDS* (Kabsch, 1988), *d\*TREK* (Pflugrath, 1999) and *XGEN* (A. J. Howard; <http://xgen.iit.edu/>).

The optimal strategy of diffraction data collection may vary considerably depending on many factors, such as the unit-cell parameters and symmetry, the crystal orientation on the goniostat, its mosaicity, susceptibility to radiation damage and diffraction resolution limit, as well as on the purpose for which

the data are intended (Dauter, 1999). Several strategy programs are available and all popular data-processing systems have such routines built in. Perhaps the most comprehensive and useful is the program *BEST* (Popov & Bourenkov, 2003), which apart from geometric considerations gives estimations of statistics related to the observed intensities. At present, strategy decisions for a particular case have to be made by the experimenter, but it may be expected that newer versions of the beamline/data-processing software will have 'intelligent' strategy routines that are able to automatically provide the set of optimal data-collection parameters or even be able to adjust such parameters on-the-fly during the diffraction experiment.

Efforts are under way to extend and link the data-processing software with other programs to include control of the data-acquisition process and even subsequent steps of the crystal structure solution, *e.g.* in the BlueIce/DCS system (McPhillips *et al.*, 2002). It is expected that in the not too distant future the whole macromolecular structure analysis from diffraction data acquisition to final model refinement will be performed at the data-collection facility while the crystal is still available for further data collection, if required.

### 2.3. Phasing

The solution of a novel crystal structure, as 50 years ago, requires the presence of heavy or anomalously scattering atoms to provide initial phase estimations. The phasing methods are therefore still based on the general principles developed for isomorphous replacement, including the anomalous component. However, particular realisations of phasing procedures have evolved very significantly and are much more powerful than those of a few years ago. In the past, when intensity estimations were in general not very accurate and tunable X-ray sources were not available, the most popular method for phase determination was the multiple isomorphous replacement (MIR) approach, with the anomalous signal treated as an auxiliary source of phasing.

With the availability of data-collection techniques able to provide very accurately measured intensities, emphasis shifted towards the use of the anomalous signal as a primary source of phase estimations through the multiple or single-wavelength anomalous diffraction approaches (MAD or SAD). The possibility of introducing selenomethionine into proteins (Hendrickson *et al.*, 1990) resulted in the MAD phasing vehicle dominating in recent years (Hendrickson, 1999). The potential anomalous scatterers can be easily identified by the recently introduced PIXE (proton-induced X-ray emission) technique (Garman, 1999). The initial analytical approach to MAD (Karle, 1989; Hendrickson, 1991) has been superseded by treating MAD as a special case of MIRAS (Ramakrishnan & Biou, 1997) and now is usually treated by more elaborate statistical approaches, *e.g.* in *SHARP* (de La Fortelle & Bricogne, 1997) and *PHASER* (Storoni *et al.*, 2004).

Single-wavelength phasing, pioneered by Hendrickson & Teeter (1981) and Wang (1985), is technically simpler than

MAD but requires accurate estimation of anomalous intensity differences. With recent progress in data-collection techniques and the current trend towards high-throughput structure determination, the SAD approach has acquired increasing popularity and favorably competes with MAD (Rice *et al.*, 2000; Dauter *et al.*, 2002). David Blow in one of his last papers stated that 'MAD is giving way to SAD' (Blow, 2003). Indeed, at Argonne's SBC ID19, one of the most productive synchrotron beamlines in the world, the majority of novel structures are currently solved by SAD, even when crystals contain selenomethionine (A. Joachimiak, private communication). SAD may need more accurate intensity estimations than MAD, especially when weak anomalous scatterers such as phosphorus or sulfur are used (Dauter & Adamiak, 2001; Ramagopal *et al.*, 2003), but does not require precise wavelength tuning and can be performed in home laboratories with Cu or Cr X-ray sources. The introduction of the chromium rotating anode by Rigaku/MSU (Yang *et al.*, 2003) and designs of new synchrotron beamlines optimized for the use of long-wavelength radiation (*e.g.* at Diamond; <http://www.diamond.ac.uk/Activity/Beamlines/>) are deliberately intended for the application of SAD phasing.

The theoretical basis of heavy-atom phasing (Green *et al.*, 1954; Blow & Crick, 1959; Rossmann, 1961; Matthews, 1966) was extended to implement more statistically valid approaches based on maximum likelihood and a number of powerful and general programs are currently in wide use: *MLPHARE* (Otwinowski, 1991), *SHARP* (de La Fortelle & Bricogne, 1997), *SOLVE* (Terwilliger & Berendzen, 1999), *PHASES* (Furey & Swaminathan, 1997), *SIR2004* (Burla *et al.*, 2005) and *CRANK* (Ness *et al.*, 2004). Improvements in phasing routines are making them more powerful and applicable to more difficult cases. A recent modification in *SHARP* permits one to take into account the effects of radiation damage on reflection intensities (Schiltz *et al.*, 2004), which is very important for proper treatment of data collected at the strongest synchrotron beamlines.

Molecular replacement (MR), originally initiated by Rossmann (1972), is currently responsible for about half of all structures deposited in the PDB, which is obviously a consequence of the availability of many different models with various protein folds. This approach has also evolved significantly in recent years. Traditionally in molecular replacement, the orientation of the molecule and its positioning have been addressed separately by first performing the three-dimensional rotational search followed by the three-dimensional translation function, as programmed in *AMoRe* (Navaza, 1994) and *MOLREP* (Vagin & Teplyakov, 1997). Newer programs performing the six-dimensional search (*EPMR*, Kissinger *et al.*, 1999; *Queen of Spades*, Glykos & Kokkinidis, 2001) and implementing maximum-likelihood principles (*PHASER*; Storoni *et al.*, 2004) are more powerful and more successful in difficult cases with less similar search models. It may be expected that in the future it will be possible to solve protein structures starting with small fragments of a typical helix or sheet.

Apart from the workhorses of protein crystallographic phasing, that is approaches based on the presence of heavy/anomalous atoms (various versions of isomorphous replacement) or of known similar structural models (molecular replacement), several other methods are being pursued. Direct methods, utilizing exclusively native intensities, are responsible for the solution of almost all small organic and inorganic structures. However, their use in macromolecular crystallography is limited, since they require diffraction data extending to atomic resolution beyond 1.2 Å (Sheldrick, 1990; Morris *et al.*, 2004), although the presence of heavy atoms seems to relax this requirement to some extent. Several programs, some of which initially were intended for small-molecule crystallography, have been generalized and successfully used in macromolecular crystallography, not only for the location of heavy or anomalous atoms, but also for solution of atomic resolution structures. The most powerful ones are based on the dual direct- and reciprocal-space recycling principle, as implemented in *SnB* (Miller *et al.*, 1994), *SHELXD* (Sheldrick & Gould, 1995) or *ACORN* (Foadi *et al.*, 2000). *CRUNCH* (de Graaff *et al.*, 2001) uses another approach through matrix methods.

The three-beam phasing method (Hümmer *et al.*, 1991) offers a truly direct source of experimental phases (or rather their invariant combinations). It is not yet used in practice for solving new structures, but further development continues at several synchrotron facilities (Weckert & Hümmer, 1997) equipped with multi-circle diffractometers. It has been also proposed as a validation tool for refined phases (Soares *et al.*, 2003).

A novel method of phasing (or imaging) applicable to very large structures in the crystalline or even the non-crystalline state in conjunction with superbright X-ray sources such as free-electron lasers is the method based on measuring diffracted intensities between Bragg reflections (Miao *et al.*, 1998, 2001). This approach is being developed by, among others, David Sayre, one of the founders of direct methods 50 years ago (Sayre, 1952).

In the early 1980s, B.-C. Wang proposed solvent flattening as a powerful method of phase improvement and extension (Wang, 1985). It is based on the existence of two different regions within protein crystals, the protein and solvent, where the noise in the solvent region of the electron-density map is iteratively filtered out. Other restraints based on the known expected behavior of the electron density are also very powerful, such as the use of non-crystallographic symmetry averaging (Bricogne, 1976) and histogram matching (Zhang & Main, 1990). These density-modification procedures have been incorporated into many programs, *e.g.* *DM* (Cowtan, 1994), *RESOLVE* (Terwilliger, 2000a) and *SOLOMON* (Abrahams & Leslie, 1996), and are routinely used in current practices.

#### 2.4. Model building and refinement

Early model building, before the era of computer graphics, involved the construction of structures from pieces of metal

wire or, later, from preformed plastic fragments representing rigid groups of atoms. When interactive computer graphics were introduced, the protein models were first built by fitting individual atoms into the displayed electron-density maps. The first widely used graphics program *FRODO* (Jones, 1978) already had many useful options of protein-chain tracing, manipulating groups of atoms with stereochemical restraints, introducing larger pieces from a library of typical structures *etc.* The sophistication of graphics software constantly increases (and the capability of computers also grows) and currently used programs perform many tasks in a semi-automatic fashion, where the human operator often needs only to check visually the correctness of the resulting action. The most widely used graphics systems are *O* (Jones *et al.*, 1991), *XtalView* (McRee, 1999), *QUANTA* (Accelrys Inc., San Diego; Oldfield, 2000), *COOT* (Emsley & Cowtan, 2004) and *MAIN* (Turk, 2000, 2004).

In fact, graphics programs are today used mainly to check the correctness of the model and its fit to density or to modify selected fragments of the structure, since the main task of the primary construction of the protein chain is very often performed without human intervention by automatic model-building programs such as *ARP/wARP* (Perrakis *et al.*, 1999), *RESOLVE* (Terwilliger, 2000a) and *MAID* (Levitt, 2001). These programs evolve constantly and become more powerful and capable of tracing correct models in low-resolution maps.

*ARP/wARP* performs model building jointly with refinement and this tendency towards integration of tasks which traditionally used to be executed sequentially is clearly beneficial and will certainly will be continued, *e.g.* refinement with concomitant validation, model display with real-space fitting and refinement *etc.*

Models of early protein X-ray structures were not refined at all. The first attempts at refinement used the difference Fourier technique (Diamond, 1971), where shifts were applied to atoms on the basis of the neighboring peaks in the  $F_o - F_c$  map. Least-squares refinement was applied for the first time in refining the structure of rubredoxin by the Seattle group (Watenpaugh *et al.*, 1972, 1980). The least-squares approach interspersed with geometry regularization was introduced later (Agarwal, 1978; Isaacs & Agarwal, 1978), followed by the least-squares method with built-in geometry restraints realised in *PROLSQ* (Konnert, 1976; Konnert & Hendrickson, 1980) and other programs (*TNT*; Tronrud *et al.*, 1987). Least-squares refinement with geometry restraints expressed as energy terms was programmed in *X-PLOR* (Brünger *et al.*, 1987) and *CNS* (Brünger *et al.*, 1998). *SHELXL* (Sheldrick & Schneider, 1997) evolved from small-molecule crystallography and has many options not available in other refinement programs. The newer refinement software systems, such as *REFMAC* (Murshudov *et al.*, 1997) or *BUSTER* (Bricogne & Irwin, 1996), are based on maximum-likelihood principles. All these programs constantly evolve, becoming ever more powerful.

Many important improvements have been incorporated in refinement programs. Among them are the validation tool  $R_{\text{free}}$  (Brünger, 1992), estimation of the positional uncertainties (Cruickshank, 1999), support of merohedral twinning (in

*SHELXL* and *CNS*), TLS rigid fragment refinement (in *REFMAC5*; Winn *et al.*, 2001) and torsion-angle restraints (in *CNS*). Many options introduced into macromolecular refinement programs have been adopted from small-molecule crystallographic practice and indeed the border between small and macromolecular crystallography has become fuzzy in many respects.

### 2.5. Validation

Validation of the correctness of the model of a macromolecule obtained from the crystal structure analysis is necessary for several reasons. The number of atoms in the model and therefore the number of refined parameters may not be significantly smaller than the number of measured unique reflections, especially at low resolution. Moreover, a large percentage of reflection intensities may be rather weak in relation to their estimated uncertainties. In addition, some parts of the model are usually more flexible or disordered and their conformation cannot be confidently defined by the appearance of the electron-density map. When the model is refined without restraints, even at atomic resolution, these weakly defined parts simply explode, in contrast to the well defined fragments (Dauter *et al.*, 1992; Kleywegt, 2000).

The solvent structure always presents a problem: the border region between well ordered water sites and completely featureless bulk solvent is never clear or easy to model satisfactorily. It contains partially occupied sites with high displacement parameters and may lie close to the protein surface at low resolution or contain several shells of well behaving water sites, but such a 'debatable' region always exists.

Information serving for validation of protein structures cannot be used as prior knowledge in refinement (Dodson *et al.*, 1996, 1998; Kleywegt & Jones, 1995); otherwise, the procedure would be equivalent to Baron Münchhausen (Raspe, 1786) pulling himself by his own hair from the mud. Since practically all protein models are refined with geometrical restraints, such as the classic set of bond lengths and angles formulated by Engh & Huber (1991), agreement with those parameters (often quoted as a criterion of a 'good' structure) says nothing about the correctness of the model, but only informs us about how strong a weight has been put on the geometric terms in the refinement. In fact, models refined at very high resolution show some variability of protein geometric characteristics and suggest that the protein stereochemistry should not be restrained excessively (EU 3D Validation Network, 1998).

Information not used for the creation of the model is very valuable for its validation. The  $R_{\text{free}}$  value (Brünger, 1992), which is based on reflections not contributing to the refinement, is much more informative and provides a more objective quality criterion than the standard  $R$  value based on all reflections, which can be easily abused (Kleywegt & Jones, 1995). However,  $R_{\text{free}}$  is a global parameter and should not be used for validation of individual partially occupied solvent sites (Dodson *et al.*, 1996). Similarly, since the torsion angles

are usually not restrained during refinement, their agreement with expected values in the form of the Ramachandran plot (Ramakrishnan & Ramachandran, 1965) or clustering of side-chain rotamers is extremely useful for the purpose of model validation. Several programs have been especially developed for checking the correctness of protein models, of which the most popular and comprehensive are *PROCHECK* (Laskowski *et al.*, 1993), *WHAT\_CHECK* (Hooft *et al.*, 1996), *SFCHECK* (Vaguine *et al.*, 1999) and *MOLPROBITY* (Davis *et al.*, 2004). In fact, many validation tools are built into all contemporary refinement and graphics display programs. All crystallographers can easily check the quality of their 'product' and there is no excuse for neglecting this opportunity.

### 3. Radiation damage

Although the effects of radiation damage inflicted on crystals of macromolecules were observed very early on (Blake & Phillips, 1962) and studied theoretically many years ago (Hendrickson, 1976; Henderson, 1990), a revival of interest in this phenomenon has occurred in the last few years (see the November 2002 and May 2005 issues of the *Journal of Synchrotron Radiation*). The routine use of very bright synchrotron beamlines has shown that serious radiation damage can occur after a few minutes of irradiation of crystals, even if they are cryocooled (Burmeister, 2000; Weik *et al.*, 2000; Ravelli & McSweeney, 2000). In the presence of significant radiation damage, the increased redundancy of measurements may in effect lead to the deterioration of the overall data quality instead of its improvement.

The first notable effect of radiation damage is non-isomorphism between successive batches of data measured from the same crystal. This may seriously impair the phasing signal expected from the data successively collected at different wavelengths within a MAD experiment. The observed non-isomorphism is caused by specific structural changes, such as the breakage of disulfide bonds, decarboxylation of acids (Weik *et al.*, 2000) or rupture of the covalent bonds to heavier atoms, such as C—Br (Ravelli *et al.*, 2003), C—I (Zwart *et al.*, 2004) or S—Hg (Ramagopal *et al.*, 2005), and many smaller changes within the protein and solvent regions. All macromolecular crystals are susceptible to radiation damage to various extents, although the presence in the solvent of certain scavengers of radicals, such as, for example, ascorbic acid (Murray & Garman, 2002), can partially diminish this effect.

However, if radiation damage is diagnosed properly, its effects can be to some extent diminished or even used beneficially. If the data set is sufficiently redundant, the intensities of individual reflections can be extrapolated to zero irradiation dose (Diederichs *et al.*, 2003). The isomorphous intensity differences between the early and late parts of the data can then successfully be used for phasing (Ravelli *et al.*, 2003; Banumathi *et al.*, 2004; Schiltz *et al.*, 2004; Ramagopal *et al.*, 2005). Research on the effects of radiation damage is progressing and certainly in the future this effect will be more

satisfactorily taken into account in various crystallographic data-reduction, phasing and refinement programs.

#### 4. Automation

As mentioned in the previous sections, significant efforts have been directed in recent years towards the automation of practically all steps of protein structure elucidation (Abola *et al.*, 2000). As an example, the researcher may consider the High-Throughput Protein Crystallography (HTPX) initiative, a joint project of several laboratories in the UK aimed at the unification of all procedures of protein structure determination into a single all-encompassing interface to be used locally or remotely, <http://www.e-htpx.ac.uk/>, or activities at various synchrotron facilities (Leonard *et al.*, 2005; Panjikar *et al.*, 2005).

A very important role in this trend is played by the support, financial and psychological, of the high-throughput structural genomics initiatives aimed at the rapid solution of a large number of pre-selected gene products and the creation of a comprehensive catalogue of protein folds. It seems that at present, after the first few years of structural genomics projects, the highest benefit of these generously supported activities lies in the development of crystallographic methods, rather than in the number of the elucidated new protein folds. However, PX methodology has matured considerably to such an extent that it may be soon expected that most protein crystal structures will be solved in an almost completely automatic way. This may be sad news for more 'traditionally' inclined crystallographers, but is definitely good news for biologists interested in analyzing the functional consequences of macromolecular structures without learning the 'ins and outs' of the crystallographic trade. The possibility of solving protein structures automatically, with minimum human intervention, is certainly important for researchers in pharmaceutical companies.

Obviously, not all crystals will be susceptible to fully automatic approaches. There will always be some unusually difficult structures or crystals displaying various kinds of pathological behavior (Dauter *et al.*, 2005) and these cases will give opportunities for the specialists to 'engage their brains'.

Automation not only speeds up the process of solution of crystal structures, but also permits feedback between various stages of the process, eventually leading to better, more accurate atomic models. Inspection of the PDB statistics (O'Toole *et al.*, 2004) shows that the average quality of structures deposited from the structural genomics centers does not differ from the quality of structures elucidated in a more traditional way.

Whereas automation of the crystal structure solution process may be viewed by traditional crystallographers as the end of the heroic era, it is certainly a welcome development for biologists who want to use the knowledge of a protein structure to attack various biological problems. However, it would be advantageous if the automatic 'black boxes' had some built-in red lights with warnings: 'data completeness too

low', 'too many overloaded reflections', 'crystal highly twinned', 'too much radiation damage for MAD' *etc.*

#### 5. Conclusions

A multitude of protein structures and particularly their complexes with various ligands or other proteins from very many biological sources still remains to be solved. This will certainly be accomplished in the future by using the existing, perhaps more sophisticated and more automatic, approaches of protein crystallography. However, larger complexes and molecular machines will not be able to produce large well ordered crystals, only possibly small ones. Therefore, they will require very intense X-ray sources, such as the FEL, to be investigated by X-ray diffraction methods. It can be expected that the forefront of PX will evolve towards this direction in the future (Harrison, 2004), whereas crystal structures of individual proteins will be solved by the highly automatic methods currently being developed.

According to Stevens (2004), these two approaches to the solution of protein structures can be classified as systems-oriented and discovery-oriented. The system-oriented approach corresponds to the traditional, slower, but more focused way of solving crystal structures of macromolecules and the discovery-oriented approach is, in other words, the structural genomics, speedier and wider track. It seems that at least in the near future macromolecular crystallography will move along these two parallel paths.

#### References

- Abad-Zapatero, C. (2004). *Structure*, **12**, 523–527.  
 Abergel, C. (2004). *Acta Cryst.* **D60**, 1413–1416.  
 Abola, E., Kuhn, P., Earnest, T. & Stevens, R. C. (2000). *Nature Struct. Biol.* **7**, 973–977.  
 Abraham, E. P. & Robinson, R. (1937). *Nature (London)*, **140**, 24.  
 Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* **D52**, 30–42.  
 Agarwal, R. C. (1978). *Acta Cryst.* **A34**, 791–809.  
 Alderton, G., Ward, W. H. & Fevold, J. (1945). *J. Biol. Chem.* **157**, 43–58.  
 Arndt, U. W. (2003). *Methods Enzymol.* **368**, 21–42.  
 Arndt, U. W., Duncumb, P., Long, J. V. P., Pina, L. & Inneman, A. (1998). *J. Appl. Cryst.* **31**, 733–741.  
 Arndt, U. W., Long, J. V. P. & Duncumb, P. (1998). *J. Appl. Cryst.* **31**, 936–944.  
 Arndt, U. W. & Phillips, D. C. (1961). *Acta Cryst.* **14**, 807–818.  
 Arndt, U. W. & Willis, B. T. M. (1966). *Single-Crystal Diffractometry*. Cambridge University Press.  
 Arndt, U. W. & Wonacott, A. J. (1977). *The Rotation Method in Crystallography*. Amsterdam: North Holland.  
 Banumathi, S., Zwart, P. H., Ramagopal, U. A., Dauter, M. & Dauter, Z. (2004). *Acta Cryst.* **D60**, 1085–1093.  
 Bergfors, T. M. (1999). *Protein Crystallization: Techniques, Strategies and Tips*. La Jolla: International University Line.  
 Bernal, J. D. & Crowfoot, D. C. (1934). *Nature (London)*, **133**, 794–795.  
 Blake, C. C. F., Fenn, R. H., Johnson, L. N., Koenig, D. F., Mair, G. A., North, A. C. T., Oldham, J. W. H., Phillips, D. C., Poljak, R. J., Sarma, V. R. & Vernon, C. A. (2001). *International Tables for Crystallography*, Vol. F, edited by M. G. Rossmann & E. Arnold, pp. 745–772. Dordrecht: Kluwer Academic Publishers.

- Blake, C. C. F. & Phillips, D. C. (1962). *Biological Effects of Ionizing Radiation at the Molecular Level. International Atomic Energy Agency Symposium, Brno, Czechoslovakia, 2–6 July 1962*, pp. 183–191. Vienna: IAEA.
- Blessing, R. H. (1995). *Acta Cryst.* **A51**, 33–38.
- Blow, D. (2003). *Methods Enzymol.* **374**, 3–22.
- Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.
- Bricogne, G. (1976). *Acta Cryst.* **A32**, 832–847.
- Bricogne, G. & Irwin, J. J. (1996). *Proceedings of the CCP4 Study Weekend. Macromolecular Refinement*, edited by E. Dodson, M. Moore, A. Ralph & S. Bailey, pp. 85–92. Warrington: Daresbury Laboratory.
- Brönnimann, C., Bühler, C., Eikenberry, E. F., Horisberger, R., Hülsen, G., Schmitt, B., Schulze-Briese, C., Suzuki, M., Tomizaki, T., Toyokawa, H. & Wagner, A. (2004). *Synchrotron Radiat. News*, **17**, 23–30.
- Brünger, A. T. (1992). *Nature (London)*, **355**, 472–474.
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* **D54**, 905–921.
- Brünger, A. T., Kuriyan, J. & Karplus, M. (1987). *Science*, **235**, 458–460.
- Buerger, M. J. (1964). *The Precession Method*. New York: Wiley.
- Burla, M. C., Caliandro, R., Camalli, M., Carrozzini, B., Casciarano, G. L., De Caro, L., Giacovazzo, C., Polidori, G. & Spagna, T. (2005). *J. Appl. Cryst.* **38**, 381–388.
- Burmeister, W. P. (2000). *Acta Cryst.* **D56**, 328–341.
- Chayen, N. E., Shaw Stewart, P. D. & Blow, D. M. (1992). *J. Cryst. Growth*, **168**, 175–180.
- Chayen, N. E., Shaw Stewart, P. D., Maeder, D. L. & Blow, D. M. (1990). *J. Appl. Cryst.* **23**, 297–302.
- Chiu, M. L., Nollert, P., Loewen, M. C., Belrhali, H., Pebay-Peyroula, E., Rosenbusch, J. P. & Landau, E. M. (2000). *Acta Cryst.* **D56**, 781–784.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Cowtan, K. (1994). *Jnt CCP4/ESF-EACMB Newsl. Protein Crystallogr.* **31**, 34–38.
- Cruickshank, D. W. J. (1999). *Acta Cryst.* **D55**, 583–601.
- Dauter, Z. (1999). *Acta Cryst.* **D55**, 1703–1717.
- Dauter, Z. & Adams, D. A. (2001). *Acta Cryst.* **D57**, 990–995.
- Dauter, Z., Botos, I., La Ronde-Le Blanc, N. & Wlodawer, A. (2005). *Acta Cryst.* **D61**, 967–975.
- Dauter, Z., Dauter, M. & Dodson, E. J. (2002). *Acta Cryst.* **D58**, 494–506.
- Dauter, Z., Sieker, L. C. & Wilson, K. S. (1992). *Acta Cryst.* **B48**, 42–59.
- Davies, D. R. & Segal, D. M. (1971). *Methods Enzymol.* **22**, 266–269.
- Davis, I. W., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2004). *Nucleic Acids Res.* **32**, W615–W619.
- Diamond, R. (1971). *Acta Cryst.* **A27**, 436–452.
- Diederichs, K., McSweeney, S. & Ravelli, R. G. B. (2003). *Acta Cryst.* **D59**, 903–909.
- Dodson, E. J., Davies, G. J., Lamzin, V. S., Murshudov, G. N. & Wilson, K. S. (1998). *Structure*, **6**, 685–690.
- Dodson, E. J., Kleywegt, G. J. & Wilson, K. S. (1996). *Acta Cryst.* **D52**, 228–234.
- Ducruix, A. & Giegé, R. (1992). *Crystallization of Nucleic Acids and Proteins. A Practical Approach*. Oxford University Press.
- Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* **D60**, 2126–2132.
- Engh, R. & Huber, R. (1991). *Acta Cryst.* **A47**, 392–400.
- EU 3D Validation Network (1998). *J. Mol. Biol.* **276**, 417–436.
- Foadi, J., Woolfson, M. M., Dodson, E. J., Wilson, K. S., Jia-xing, Y. & Chao-de, Z. (2000). *Acta Cryst.* **D56**, 1137–1147.
- Furey, W. & Swaminathan, S. (1997). *Methods Enzymol.* **277**, 590–620.
- Garavito, R. M. & Rosenbusch, J. P. (1980). *J. Cell Biol.* **86**, 327–329.
- García-Ruiz, J. M. (2003). *Methods Enzymol.* **368**, 130–154.
- García-Ruiz, J. M. & Moreno, A. (1994). *Acta Cryst.* **D50**, 484–490.
- Garman, E. F. (1999). *Structure*, **7**, R291–R299.
- Garman, E. F. & Schneider, T. R. (1997). *J. Appl. Cryst.* **30**, 211–237.
- Glykos, N. M. & Kokkinidis, M. (2001). *Acta Cryst.* **D57**, 1462–1473.
- Graaff, R. A. G. de, Hilge, M., van der Plaas, J. L. & Abrahams, J. P. (2001). *Acta Cryst.* **D57**, 1857–1862.
- Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). *Proc. R. Soc. London A*, **225**, 287–307.
- Harrison, S. (2004). *Nature Struct. Mol. Biol.* **11**, 12–15.
- Heinemann, U. (2000). *Nature Struct. Biol.* **7**, Suppl., 940–942.
- Henderson, R. (1990). *Proc. R. Soc. London B*, **241**, 6–8.
- Hendrickson, W. A. (1976). *J. Mol. Biol.* **106**, 889–893.
- Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
- Hendrickson, W. A. (1999). *J. Synchrotron Rad.* **6**, 845–851.
- Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
- Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
- Holmes, K. C. & Rosenbaum, G. (1998). *J. Synchrotron Rad.* **5**, 147–153.
- Hooft, R. W. W., Vriend, G., Sander, C. & Abola, E. E. (1996). *Nature (London)*, **381**, 272.
- Hope, H. (1988). *Acta Cryst.* **B44**, 22–26.
- Huldt, G., Szoke, A. & Hajdu, J. (2003). *J. Struct. Biol.* **144**, 219–227.
- Hümmer, K., Schwegle, W. & Weckert, E. (1991). *Acta Cryst.* **A47**, 60–62.
- Isaacs, N. W. & Agarwal, R. C. (1978). *Acta Cryst.* **A34**, 782–791.
- Jacquemet, L., Ohana, J., Joly, J., Borel, F., Pirocchi, M., Charrault, P., Bertoni, A., Israel-Gouy, P., Carpentier, P., Kozielski, F., Blot, D. & Ferrer, J.-L. (2004). *Structure*, **12**, 1219–1225.
- Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.
- Jiang, J. & Sweet, R. M. (2004). *J. Synchrotron Rad.* **11**, 319–327.
- Jones, T. A. (1978). *J. Appl. Cryst.* **11**, 268–272.
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* **A47**, 110–119.
- Judson, H. F. (1979). *The Eighth Day of Creation*. New York: Simon & Schuster.
- Kabsch, W. (1988). *J. Appl. Cryst.* **21**, 916–924.
- Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
- Karle, J. (1989). *Int. J. Quant. Chem. Quant. Biol. Symp.* **7**, 357–367.
- Kendrew, J. C. (1963). *Science*, **139**, 1259–1266.
- Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C. & Shore, V. C. (1960). *Nature (London)*, **185**, 422–427.
- Kiefersauer, R., Than, M. E., Dobbek, H., Gremer, L., Melero, M., Strobl, S., Dias, J. M., Soulimane, T. & Huber, R. (2000). *J. Appl. Cryst.* **33**, 1223–1230.
- Kissinger, C. R., Gehlhaar, D. K. & Fogel, D. B. (1999). *Acta Cryst.* **D55**, 484–491.
- Kitago, Y., Watanabe, N. & Tanaka, I. (2005). *Acta Cryst.* **D61**, 1013–1021.
- Kleywegt, G. J. (2000). *Acta Cryst.* **D56**, 249–265.
- Kleywegt, G. J. & Jones, T. A. (1995). *Structure*, **3**, 535–540.
- Konnert, J. H. (1976). *Acta Cryst.* **A32**, 614–617.
- Konnert, J. H. & Hendrickson, W. A. (1980). *Acta Cryst.* **A36**, 344–350.
- La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
- Landau, E. M. & Rosenbusch, J. P. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 14532–14535.
- Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
- Leonard, G. A., Sainz, G., de Backer, M. M. E. & McSweeney, S. (2005). *Acta Cryst.* **D61**, 388–396.
- Leslie, A. G. W. (1992). *Jnt CCP4/ESF-EACBM Newsl. Protein Crystallogr.* **26**.
- Levitt, D. G. (2001). *Acta Cryst.* **D57**, 1013–1019.

- McPherson, A. (1982). *Preparation and Analysis of Protein Crystals*. New York: Wiley.
- McPhillips, T. M., McPhillips, S. E., Chiu, H.-J., Cohen, A. E., Deacon, A. M., Ellis, P. J., Garman, E., Gonzalez, A., Sauter, N. K., Phizackerley, R. P., Soltis, S. M. & Kuhn, P. (2002). *J. Synchrotron Rad.* **9**, 401–406.
- McRee, D. (1999). *J. Struct. Biol.* **125**, 156–165.
- Matthews, B. W. (1966). *Acta Cryst.* **20**, 82–86.
- Matthews, B. W. (2003). *Methods Enzymol.* **368**, 3–11.
- Miao, J., Hodgson, K. O. & Sayre, D. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 6641–6645.
- Miao, J., Sayre, D. & Chapman, H. N. (1998). *J. Opt. Soc. Am. A*, **15**, 1662–1669.
- Michel, H. & Oesterhelt, D. (1980). *Proc. Natl Acad. Sci. USA*, **77**, 1283–1285.
- Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.
- Morris, R. J., Blanc, E. & Bricogne, G. (2004). *Acta Cryst.* **D60**, 227–240.
- Murray, J. & Garman, E. (2002). *J. Synchrotron Rad.* **9**, 347–354.
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* **D53**, 240–255.
- Navaza, J. (1994). *Acta Cryst.* **A50**, 157–163.
- Ness, S. R., de Graaff, R. A. G., Abrahams, J. P. & Pannu, N. S. (2004). *Structure*, **12**, 1753–1761.
- Ng, J. D., Gavira, J. A. & García-Ruiz, J. M. (2003). *J. Struct. Biol.* **142**, 218–231.
- Northrop, J. H. (1929). *Science*, **69**, 580.
- Northrop, J. H. (1930). *J. Gen. Physiol.* **13**, 739–766.
- Northrop, J. H. (1932). *J. Gen. Physiol.* **16**, 267–294.
- Oldfield, T. J. (2000). *Acta Cryst.* **A56**, s27.
- O'Toole, N., Grabowski, M., Otwinowski, Z., Minor, W. & Cygler, M. (2004). *Proteins*, **56**, 201–210.
- Otwinowski, Z. (1991). *Proceedings of the CCP4 Study Weekend. Isomorphous Replacement and Anomalous Scattering*, edited by W. Wolf, P. R. Evans & A. G. W. Leslie, pp. 80–86. Warrington: Daresbury Laboratory.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2005). *Acta Cryst.* **D61**, 449–457.
- Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
- Perutz, M. F. (1963). *Science*, **140**, 863–869.
- Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G. & North, A. C. T. (1960). *Nature (London)*, **185**, 416–422.
- Pflugrath, J. W. (1999). *Acta Cryst.* **D55**, 1718–1725.
- Phillips, J. C., Wlodawer, W., Yevitz, M. M. & Hodgson, K. O. (1976). *Proc. Natl Acad. Sci. USA*, **73**, 128–132.
- Popov, A. N. & Bourenkov, G. P. (2003). *Acta Cryst.* **D59**, 1145–1153.
- Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Acta Cryst.* **D59**, 1020–1027.
- Ramagopal, U. A., Thirumurugan, R., Fedorova, E., Dauter, Z. & Almo, S. C. (2005). *Acta Cryst.* **D61**, 1289–1298.
- Ramakrishnan, V. & Biou, V. (1997). *Methods Enzymol.* **276**, 538–557.
- Ramakrishnan, C. & Ramachandran, G. N. (1965). *Biophys. J.* **5**, 909–933.
- Raspe, R. E. (1786). *The Surprising Adventures of Baron Münchhausen*. London: Kearsley.
- Ravelli, R. B. G. & McSweeney, S. M. (2000). *Structure*, **8**, 315–328.
- Ravelli, R. B. G., Schröder-Leiros, H.-K., Pan, B., Caffrey, M. & McSweeney, S. M. (2003). *Structure*, **11**, 217–224.
- Rice, L. M., Earnest, T. N. & Brunger, A. T. (2000). *Acta Cryst.* **D56**, 1413–1420.
- Rosenbaum, G., Holmes, K. C. & Witz, I. (1971). *Nature (London)*, **230**, 434–437.
- Rossmann, M. G. (1961). *Acta Cryst.* **14**, 383–388.
- Rossmann, M. G. (1972). *The Molecular Replacement Method*. New York: Gordon & Breach.
- Rossmann, M. G. (2003). *Methods Enzymol.* **368**, 11–21.
- Rossmann, M. G. & van Beek, C. G. (1999). *Acta Cryst.* **D55**, 1631–1640.
- Sayre, D. (1952). *Acta Cryst.* **5**, 60–65.
- Schiltz, M., Dumas, P., Ennifar, E., Flensburg, C., Paciorek, W., Vonrhein, C. & Bricogne, G. (2004). *Acta Cryst.* **D60**, 1024–1031.
- Shakespeare, W. (1603). *Hamlet*, Act 1, Scene 5.
- Sharp, K. R. (2004). Nobugs2004 Conference, PSI, Villigen, Switzerland. <http://lns00.psi.ch/nobugs2004/>.
- Sheldrick, G. M. (1990). *Acta Cryst.* **A46**, 467–473.
- Sheldrick, G. M. & Gould, R. O. (1995). *Acta Cryst.* **B51**, 423–431.
- Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
- Soares, A. S., Caspar, D. L. D., Weckert, E., Héroux, A., Hölzer, K., Schroer, K., Zellner, J., Schneider, D., Nolan, W. & Sweet, R. M. (2003). *Acta Cryst.* **D59**, 1716–1724.
- Stanley, W. M. (1935). *Science*, **81**, 644–645.
- Stevens, R. (2004). *Nature Struct. Mol. Biol.* **11**, 293–295.
- Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). *Acta Cryst.* **D60**, 432–438.
- Sumner, J. B. (1926a). *J. Biol. Chem.* **69**, 435–441.
- Sumner, J. B. (1926b). *J. Biol. Chem.* **70**, 97–98.
- Sumner, J. B. & Dounce, A. L. (1937a). *J. Biol. Chem.* **121**, 417–424.
- Sumner, J. B. & Dounce, A. L. (1937b). *Science*, **85**, 366–367.
- Teng, T.-Y. (1990). *J. Appl. Cryst.* **23**, 387–391.
- Terwilliger, T. C. (2000a). *Acta Cryst.* **D56**, 965–972.
- Terwilliger, T. C. (2000b). *Nature Struct. Biol.* **7**, Suppl., 935–939.
- Terwilliger, T. C. & Berendzen, J. (1999). *Acta Cryst.* **D55**, 849–861.
- Tronrud, D. E., Ten Eyck, L. F. & Matthews, B. W. (1987). *Acta Cryst.* **A43**, 489–501.
- Turk, D. (2000). *Acta Cryst.* **A56**, s27.
- Turk, D. (2004). *Acta Cryst.* **A60**, s16.
- Vagin, A. & Teplyakov, A. (1997). *J. Appl. Cryst.* **30**, 1022–1025.
- Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Cryst.* **D55**, 191–205.
- Vergara, A., Lorber, B., Zagari, A. & Giegé, R. (2003). *Acta Cryst.* **D59**, 2–15.
- Walsh, M. A., Dementieva, I., Evans, G., Sanishvili, R. & Joachimiak, A. (1999). *Acta Cryst.* **D55**, 1168–1173.
- Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
- Watanabe, N., Murai, H. & Tanaka, I. (2002). *Acta Cryst.* **D58**, 1527–1530.
- Watenpaugh, K. D., Sieker, L. C., Herriott, J. R. & Jensen, L. H. (1972). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 359–367.
- Watenpaugh, K. D., Sieker, L. C. & Jensen, L. H. (1980). *J. Mol. Biol.* **138**, 615–633.
- Weckert, E. & Hümmel, K. (1997). *Acta Cryst.* **A53**, 108–143.
- Weik, M., Ravelli, R. B. G., Kryger, G., McSweeney, S., Raves, M. L., Harel, M., Gros, P., Silman, I., Kroon, J. & Sussman, J. L. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 623–628.
- Weselak, M., Patch, M. G., Selby, T. L., Knebel, G. & Stevens, R. C. (2003). *Methods Enzymol.* **368**, 45–76.
- Winn, M. D., Isupov, M. N. & Murshudov, G. N. (2001). *Acta Cryst.* **D57**, 122–133.
- Yang, C., Pflugrath, J. W., Courville, D. A., Stence, C. N. & Ferrara, J. D. (2003). *Acta Cryst.* **D59**, 1943–1957.
- Yokohama, S., Hirota, H., Kigawa, T., Yabuki, T., Shirouzu, M., Terada, T., Ito, Y., Matsuo, Y., Kuroda, Y., Nishimura, Y., Kyogoku, Y., Miki, K., Masui, R. & Kuramitsu, S. (2000). *Nature Struct. Biol.* **7**, Suppl., 943–945.
- Zeppezauer, M., Eklund, H. & Zeppezauer, E. S. (1968). *Arch. Biochem. Biophys.* **126**, 564–573.
- Zhang, K. Y. J. & Main, P. (1990). *Acta Cryst.* **A46**, 41–46.
- Zwart, P. H., Banumathi, S., Dauter, M. & Dauter, Z. (2004). *Acta Cryst.* **D60**, 1958–1963.