International Tables for Crystallography (2006). Vol. F, Chapter 1.2, pp. 4–9.

1.2. Historical background

BY M. G. ROSSMANN

1.2.1. Introduction

Crystallography ranks with astronomy as one of the oldest sciences. Crystals, in the form of precious stones and common minerals, have attractive properties on account of their symmetry and their refractive and reflective properties, which result in the undefinable quality called beauty. Natural philosophers have long pondered the unusual properties seen in the discontinuous surface morphologies of crystals. Hooke (1665) and Huygens (1690) came close to grasping the way repeating objects create discrete crystal faces with reproducible interfacial angles. The symmetry of mineral crystals was explored systematically in the 18th and 19th centuries by measuring the angles between crystal faces, leading to the classification into symmetry systems from triclinic to cubic and the construction of symmetry tables (Schoenflies, 1891; Hilton, 1903; Astbury *et al.*, 1935) – the predecessors of today's *International Tables.*

1.2.2. 1912 to the 1950s

It was not until the interpretation of the first X-ray diffraction experiments by Max von Laue and Peter Ewald in 1912 that it was possible to ascertain the size of the repeating unit in simple crystals. Lawrence Bragg, encouraged by his father, William Bragg, recast the Laue equations into the physically intuitive form now known as 'Bragg's law' (Bragg & Bragg, 1913). This set the stage for a large number of structure determinations of inorganic salts and metals. The discovery of simple structures (Bragg, 1913), such as that of NaCl, led to a good deal of acrimony, for crystals of such salts were shown to consist of a uniform distribution of positive and negative ions, rather than discrete molecules. These early structure determinations were based on trial and error (sometimes guided by the predictions of Pope and Barlow that were based on packing considerations) until a set of atomic positions could be found that satisfied the observed intensity distribution of the X-ray reflections. This gave rise to rather pessimistic estimates that structures with more than about four independent atomic parameters would not be solvable.

The gradual advance in X-ray crystallography required a systematic understanding and tabulation of space groups. Previously, only various aspects of three-dimensional symmetry operations appropriate for periodic lattices had been listed. Consequently, in 1935, the growing crystallographic community put together the first set of *Internationale Tabellen* (Hermann, 1935), containing diagrams and information on about 230 space groups. After World War II, these tables were enlarged and combined with Kathleen Lonsdale's structure-factor formulae (Lonsdale, 1936) in the form of *International Tables* Volume I (Henry & Lonsdale, 1952). Most recently, they have again been revised and extended in Volume A (Hahn, 1983).

Simple organic compounds started to be examined in the 1920s. Perhaps foremost among these is the structure of hexamethylbenzene by Kathleen Lonsdale (Lonsdale, 1928). She showed that, as had been expected, benzene had a planar hexagonal structure. Another notable achievement of crystallography was made by J. D. Bernal in the early 1930s. He was able to differentiate between a number of possible structures for steroids by studying their packing arrangements in different unit cells (Bernal, 1933). Bernal ('Sage') had an enormous impact on English crystallographers in the 1930s. His character was immortalized by the novelist C. P. Snow in his book *The Search* (Snow, 1934). By the mid-1930s, J. Monteath Robertson and I. Woodward had determined the structure of nickel phthalocyanine (Robertson, 1935) using the heavy-atom method. This was a major crystallographic success and perhaps the first time that a crystallographer had succeeded in solving a structure when little chemical information was available.

Another event which had a major impact was the determination of the absolute hand of the asymmetric carbon atom of sodium tartrate by Bijvoet (Bijvoet, 1949; Bijvoet *et al.*, 1951). By indexing the X-ray reflections with a right-handed system, he showed that the breakdown of Friedel's law in the presence of an anomalous scatterer was consistent with the asymmetric carbon atom having a hand in agreement with Fischer's convention. With that knowledge, together with the prior results of organic reaction analyses, the absolute hand of other asymmetric carbon atoms could be established. In particular, the absolute structure of naturally occurring amino acids and riboses was now determined.

Until the mid-1950s, most structure determinations were made using only projection data. This not only reduced the tremendous effort required for manual indexing and for making visual estimates of intensity measurements, but also reduced the calculation effort to almost manageable proportions in the absence of computing machines. However, the structure determination of penicillin (Crowfoot, 1948; Crowfoot *et al.*, 1949), carried out during World War II by Dorothy Hodgkin and Charles Bunn, employed some three-dimensional data. A further major achievement was the solution of the three-dimensional structure of vitamin B_{12} by Dorothy Hodgkin and her colleagues (Hodgkin *et al.*, 1957) in the 1950s. They first used a cobalt atom as a heavy atom on a vitamin B_{12} fragment and were able to recognize the 'corrin' ring structure. The remainder of the B_{12} structure was determined by an extraordinary collaboration between Dorothy Hodgkin in Oxford and Kenneth Trueblood at UCLA in Los Angeles. While Dorothy's group did the data collection and interpretation, Ken's group performed the computing on the very early electronic Standard Western Automatic Computer (SWAC). Additional help was made available by the parallel work of J. G. White at Princeton University in New Jersey. This was at a time before the internet, before e-mail, before usable transatlantic telephones and before jet travel. Transatlantic, propeller-driven air connections had started to operate only a few years earlier.

Many technical advances were made in the 1930s that contributed to the rapidly increasing achievements of crystallography. W. H. Bragg had earlier suggested (Bragg, 1915) the use of Fourier methods to analyse the periodic electron-density distribution in crystals, and this was utilized by his son, W. L. Bragg (Bragg, 1929*a*,*b*). The relationship between a Fourier synthesis and a Fourier analysis demonstrated that the central problem in structural crystallography was in the phase. Computational devices to help plot this distribution were invented by Arnold Beevers and Henry Lipson in the form of their 'Beevers–Lipson strips' (Beevers & Lipson, 1934) and by J. Monteath Robertson with his 'Robertson sorting board' (Robertson, 1936). These devices were later supplemented by the XRAC electronic analogue machine of Ray Pepinsky (Pepinsky, 1947) and mechanical analogue machines (McLachlan & Champaygne, 1946; Lipson & Cochran, 1953) until electronic digital computers came into use during the mid-1950s.

A. Lindo Patterson, inspired by his visit to England in the 1930s where he met Lawrence Bragg, Kathleen Lonsdale and J. Monteath Robertson, showed how to use F^2 Fourier syntheses for structure determinations (Patterson, 1934, 1935). When the 'Patterson' synthesis was combined with the heavy-atom method, and (later) with electronic computers, it transformed analytical organic chemistry. No longer was it necessary for teams of chemists to

labour for decades on the structure determination of natural products. Instead, a single crystallographer could solve such a structure in a period of months.

Improvements in data-collection devices have also had a major impact. Until the mid-1950s, the most common method of measuring intensities was by visual comparison of reflection 'spots' on films with a standard scale. However, the use of counters (used, for instance, by Bragg in 1912) was gradually automated and became the preferred technique in the 1960s. In addition, semi-automatic methods of measuring the optical densities along reciprocal lines on precession photographs were used extensively for early protein-structure determinations in the 1950s and 1960s.

1.2.3. The first investigations of biological macromolecules

Leeds, in the county of Yorkshire, was one of the centres of England's textile industry and home to a small research institute established to investigate the properties of natural fibres. W. T. Astbury became a member of this institute after learning about X-ray diffraction from single crystals in Bragg's laboratory. He investigated the diffraction of X-rays by wool, silk, keratin and other natural fibrous proteins. He showed that the resultant patterns could be roughly classified into two classes, α and β, and that on stretching some, for example, wool, the pattern is converted from α to β (Astbury, 1933).

Purification techniques for globular proteins were also being developed in the 1920s and 1930s, permitting J. B. Sumner at Cornell University to crystallize the first enzyme, namely urease, in 1926. Not much later, in Cambridge, J. D. Bernal and his student, Dorothy Crowfoot (Hodgkin), investigated crystals of pepsin. The resultant 1934 paper in *Nature* (Bernal & Crowfoot, 1934) is quite remarkable because of its speed of publication and because of the authors' extraordinary insight. The crystals of pepsin were found to deteriorate quickly in air when taken out of their crystallization solution and, therefore, had to be contained in a sealed capillary tube for all X-ray experiments. This form of protein-crystal mounting remained in vogue until the 1990s when crystal-freezing techniques were introduced. But, most importantly, it was recognized that the pepsin diffraction pattern implied that the protein molecules have a unique structure and that these crystals would be a vehicle for the determination of that structure to atomic resolution. This understanding of protein structure occurred at a time when proteins were widely thought to form heterogeneous micelles, a concept which persisted another 20 years until Sanger was able to determine the unique amino-acid sequences of the two chains in an insulin molecule (Sanger & Tuppy, 1951; Sanger & Thompson, 1953*a*,*b*).

Soon after Bernal and Hodgkin photographed an X-ray diffraction pattern of pepsin, Max Perutz started his historic investigation of haemoglobin.* Such investigations were, however, thought to be without hope of any success by most of the contemporary crystallographers, who avoided crystals that did not have a short (less than 4.5 Å) axis for projecting resolved atoms. Nevertheless, Perutz computed Patterson functions that suggested haemoglobin contained parallel α-keratin-like bundles of rods (Boyes-Watson *et al.*, 1947; Perutz, 1949). Perutz was correct about the α-keratin-like rods, but not about these being parallel.

In Pasadena, Pauling (Pauling & Corey, 1951; Pauling *et al.*, 1951) was building helical polypeptide models to explain Astbury's α patterns and perhaps to understand the helical structures in globular proteins, such as haemoglobin. Pauling, using his knowledge of the structure of amino acids and peptide bonds, was forced to the conclusion that there need not be an integral number of amino-acid residues per helical turn. He therefore suggested that the 'α-helix', with 3.6 residues per turn, would roughly explain Astbury's α pattern and that his proposed 'β-sheet' structure should be related to Astbury's β pattern. Perutz saw that an α-helical structure should give rise to a strong 1.5 Å-spacing reflection as a consequence of the rise per residue in an α-helix (Perutz, 1951*a*,*b*). Demonstration of this reflection in horse hair, then in fibres of polybenzyl-L-glutamate, in muscle (with Hugh Huxley) and finally in haemoglobin crystals showed that Pauling's proposed α-helix really existed in haemoglobin and presumably also in other globular proteins. Confirmation of helix-like structures came with the observation of cylindrical rods in the 6 Å-resolution structure of myoglobin in 1957 (Kendrew *et al.*, 1958) and eventually at atomic resolution with the 2 Å myoglobin structure in 1959 (Kendrew *et al.*, 1960). The first atomic resolution confirmation of Pauling's β structure did not come until 1966 with the structure determination of hen egg-white lysozyme (Blake, Mair *et al.*, 1967).

Although the stimulus for the Cochran *et al.* (1952) analysis of diffraction from helical structures came from Perutz's studies of helices in polybenzyl-L-glutamate and their presence in haemoglobin, the impact on the structure determination of nucleic acids was even more significant. The events leading to the discovery of the double-helical structure of DNA have been well chronicled (Watson, 1968; Olby, 1974; Judson, 1979). The resultant science, often known exclusively as molecular biology, has created a whole new industry. Furthermore, the molecular-modelling techniques used by Pauling in predicting the structure of α-helices and β-sheets and by Crick and Watson in determining the structure of DNA had a major effect on more traditional crystallography and the structure determinations of fibrous proteins, nucleic acids and polysaccharides.

Another major early result of profound biological significance was the demonstration by Bernal and Fankuchen in the 1930s (Bernal & Fankuchen, 1941) that tobacco mosaic virus (TMV) had a rod-like structure. This was the first occasion where it was possible to obtain a definite idea of the architecture of a virus. Many of the biological properties of TMV had been explored by Wendell Stanley working at the Rockefeller Institute in New York. He had also been able to obtain a large amount of purified virus. Although it was not possible to crystallize this virus, it was possible to obtain a diffraction pattern of the virus in a viscous solution which had been agitated to cause alignment of the virus particles. This led Jim Watson (Watson, 1954) to a simple helical structure of protein subunits. Eventually, after continuing studies by Aaron Klug, Rosalind Franklin, Ken Holmes and others, the structure was determined at atomic resolution (Holmes *et al.*, 1975), in which the helical strand of RNA was protected by the helical array of protein subunits.

1.2.4. Globular proteins in the 1950s

In 1936, Max Perutz had joined Sir Lawrence Bragg in Cambridge. Inspired in part by Keilin (Perutz, 1997), Perutz started to study crystalline haemoglobin. This work was interrupted by World War II, but once the war was over Perutz tenaciously developed a series of highly ingenious techniques. All of these procedures have their

* Perutz writes, 'I started X-ray work on haemoglobin in October 1937 and Bragg became Cavendish Professor in October 1938. Bernal was my PhD supervisor in 1937, but he had nothing to do with my choice of haemoglobin. I began this work at the suggestion of Haurowitz, the husband of my cousin Gina Perutz, who was then in Prague. The first paper on X-ray diffraction from haemoglobin (and chymotrypsin) was Bernal, Fankuchen & Perutz (Bernal *et al.*, 1938). I did the experimental work, (and) Bernal showed me how to interpret the X-ray pictures.

counterparts in modern 'protein crystallography', although few today recognize their real origin.

The first of these methods was the use of 'shrinkage' stages (Perutz, 1946; Bragg & Perutz, 1952). It had been noted by Bernal and Crowfoot (Hodgkin) in their study of pepsin that crystals of proteins deteriorate on exposure to air. Perutz examined crystals of horse haemoglobin after they were air-dried for short periods of time and then sealed in capillaries. He found that there were at least seven consecutive discrete shrinkage stages of the unit cell. He realized that each shrinkage stage permitted the sampling of the molecular transform at successive positions, thus permitting him to map the variation of the continuous transform. As he examined only the centric (*h*0*l*) reflections of the monoclinic crystals, he could observe when the sign changed from 0 to π in the centric projection (Fig. 1.2.4.1). Thus, he was able to determine the phases (signs) of the central part of the (*h*0*l*) reciprocal lattice. This technique is essentially identical to the use of diffraction data from different unit cells for averaging electron density in the 'modern' molecular replacement method. In the haemoglobin case, Patterson projections had shown that the molecules maintained their orientation relative to the *a* axis as the crystals shrank, but in the more general molecular replacement case, it is necessary to determine the relative orientations of the molecules in each cell.

The second of Perutz's techniques depended on observing changes in the intensities of low-order reflections when the concentration of the dissolved salts (*e.g.* Cs_2SO_4) in the solution between the crystallized molecules was altered (Boyes-Watson *et al.*, 1947; Perutz, 1954). The differences in structure amplitude, taken together with the previously determined signs, could then map out the parts of the crystal unit cell occupied by the haemoglobin molecule. In many respects, this procedure has its equivalent in 'solvent flattening' used extensively in 'modern' protein crystallography.

The third of Perutz's innovations was the isomorphous replacement method (Green *et al.*, 1954). The origin of the isomorphous replacement method goes back to the beginnings of X-ray crystallography when Bragg compared the diffracted intensities from crystals of NaCl and KCl. J. Monteath Robertson explored the procedure a little further in his studies of phthalocyanines. Perutz used a well known fact that dyes could be diffused into protein crystals, and, hence, heavy-atom compounds might also diffuse into and bind to specific residues in the protein. Nevertheless, the sceptics questioned whether even the heaviest atoms could make a measurable difference to the X-ray diffraction pattern of a protein.* Perutz therefore developed an instrument which quantitatively recorded the blackening caused by the reflected X-ray beam on a film. He also showed that the effect of specifically bound atoms could be observed visually on a film record of a diffraction pattern. In 1953, this resulted in a complete sign determination of the (*h*0*l*) horse haemoglobin structure amplitudes (Green *et al.*, 1954). However, not surprisingly, the projection of the molecule was not very interesting, making it necessary to extend the procedure to noncentric, three-dimensional data. It took another five years to determine the first globular protein structure to near atomic resolution.

In 1950, David Harker was awarded one million US dollars to study the structure of proteins. He worked first at the Brooklyn Polytechnic Institute in New York and later at the Roswell Park Cancer Institute in Buffalo, New York. He proposed to solve the structure of proteins on the assumption that they consisted of 'globs' which he could treat as single atoms; therefore, he could solve the structure by using his inequalities (Harker & Kasper, 1947), *i.e.*, by direct methods. He was aware of the need to use three-dimensional data, which meant a full phase determination, rather than the sign determination of two-dimensional projection data on which Perutz had concentrated. Harker therefore decided to develop automatic diffractometers, as opposed to the film methods being used at Cambridge. In 1956, he published a procedure for plotting the isomorphous data of each reflection in a simple graphical manner that allowed an easy determination of its phase (Harker, 1956). Unfortunately, the error associated with the data tended to create a lot of uncertainty.

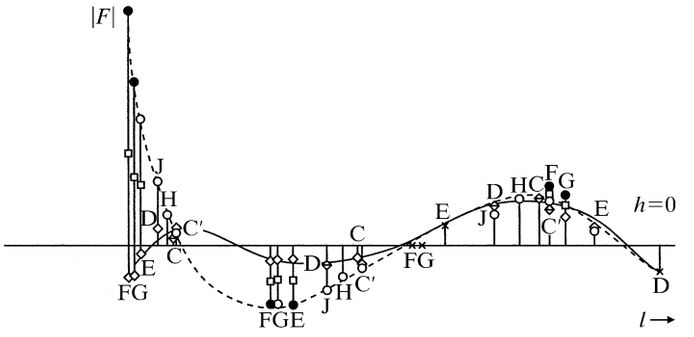

Fig. 1.2.4.1. Change of structure amplitude for horse haemoglobin as a function of salt concentration in the suspension medium of the low-order *h*0*l* reflections at various lattice shrinkage stages (C, C′, D, E, F, G, H, J). Reprinted with permission from Perutz (1954). Copyright (1954) Royal Society of London.

In the first systematic phase determination of a protein, namely that of myoglobin, phase estimates were made for about 400 reflections. In order to remove subjectivity, independent estimates were made by Kendrew and Bragg by visual inspection of the Harker diagram for each reflection. These were later compared before computing an electron-density map. This process was put onto a more objective basis by calculating phase probabilities, as described by Blow & Crick (1959) and Dickerson *et al.* (1961).

One problem with the isomorphous replacement method was the determination of accurate parameters that described the heavy-atom replacements. Centric projections were a means of directly determining the coordinates, but no satisfactory method was available to determine the relative positions of atoms in different derivatives when there were no centric projections. In particular, it was necessary to establish the relative *y* coordinates for the heavy-atom sites in the various derivatives of monoclinic myoglobin and in monoclinic horse haemoglobin. Perutz (1956) and Bragg (1958) had each proposed solutions to this problem, but these were not entirely satisfactory. Consequently, it was necessary to average the results of different methods to determine the 6 Å phases for myoglobin. However, this problem was solved satisfactorily in the structure determination of haemoglobin by using an $(F_{H1} - F_{H2})^2$ Patterson-like synthesis in which the vectors between atoms in the two heavy-atom compounds, H1 and H2, produce negative peaks (Rossmann, 1960). This technique also gave rise to the first least-squares refinement procedure to determine the parameters that define each heavy atom.

Perutz used punched cards to compute the first three-dimensional Patterson map of haemoglobin. This was a tremendous computational undertaking. However, the first digital electronic computers started to appear in the early to mid-1950s. The EDSAC1 and EDSAC2 machines were built in the Mathematical Laboratory of Cambridge University. EDSAC1 was used by John Kendrew for the

* Perutz writes, 'I measured the absolute intensity of reflexions from haemoglobin which turned out to be weaker than predicted by Wilson's statistics. This made me realise that about 99% of the scattering contributions of the light atoms are extinguished by interference and that, by contrast, the electrons of a heavy atom, being concentrated at a point, would scatter in phase and therefore make a measurable difference to the structure amplitudes.'

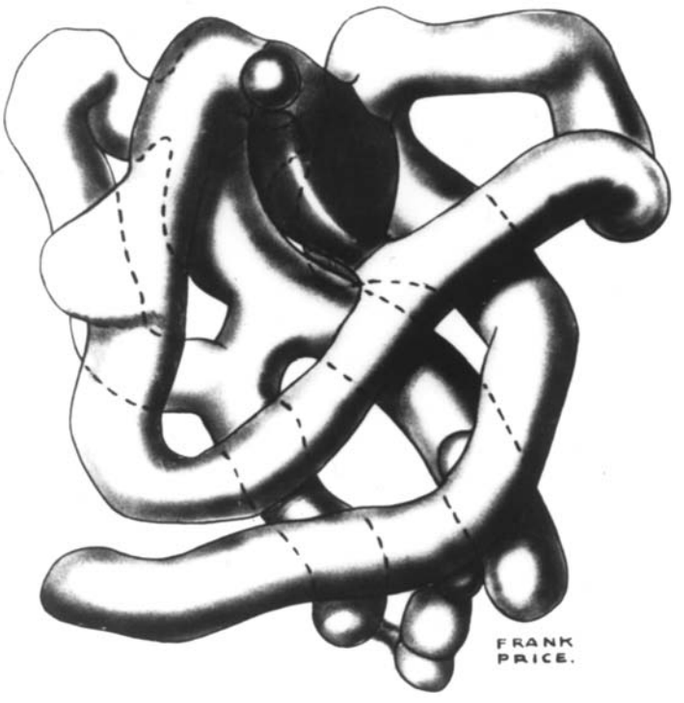

Fig. 1.2.5.1. A model of the myoglobin molecule at 6 Å resolution. Reprinted with permission from Bodo *et al.* (1959). Copyright (1959) Royal Society of London.

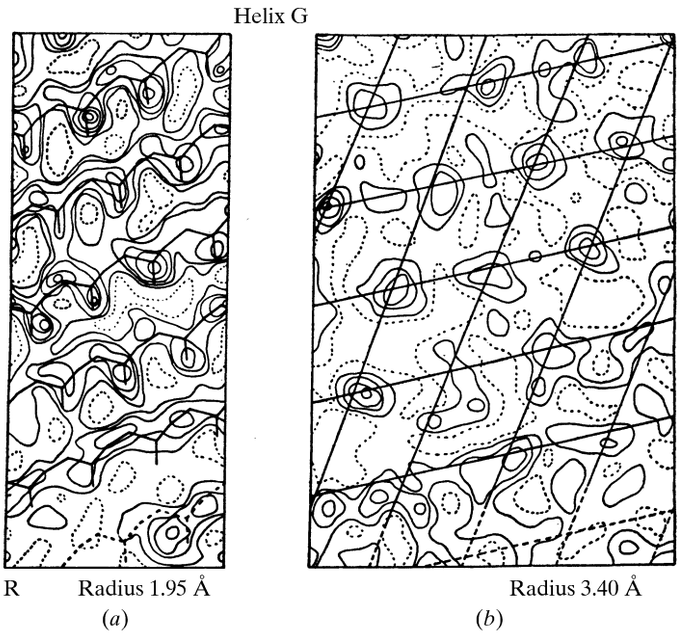

Fig. 1.2.5.2. Cylindrical sections through a helical segment of a myoglobin polypeptide chain. (*a*) The density in a cylindrical mantle of 1.95 Å radius, corresponding to the mean radius of the main-chain atoms in an α-helix. The calculated atomic positions of the α-helix are superimposed and roughly correspond to the density peaks. (*b*) The density at the radius of the β-carbon atoms; the positions of the β-carbon atoms calculated for a right-handed α-helix are marked by the superimposed grid (Kendrew & Watson, unpublished). Reprinted with permission from Perutz (1962). Copyright (1962) Elsevier Publishing Co.

6 Å-resolution map of myoglobin (Bluhm *et al.*, 1958). EDSAC2 came on-line in 1958 and was the computer on which all the calculations were made for the 5.5 Å map of haemoglobin (Cullis *et al.*, 1962) and the 2.0 Å map of myoglobin. It was the tool on which many of the now well established crystallographic techniques were initially developed. By about 1960, the home-built, one-of-a-kind machines were starting to be replaced by commercial machines. Large mainframe IBM computers (704, 709 *etc.*), together with FORTRAN as a symbolic language, became available.

1.2.5. The first protein structures (1957 to the 1970s)

By the time three-dimensional structures of proteins were being solved, Linderström-Lang (Linderström-Lang & Schellman, 1959) had introduced the concepts of 'primary', 'secondary' and 'tertiary' structures, providing a basis for the interpretation of electron-density maps. The first three-dimensional protein structure to be solved was that of myoglobin at 6 Å resolution (Fig. 1.2.5.1) in 1957 (Kendrew *et al.*, 1958). It clearly showed sausage-like features which were assumed to be α-helices. The iron-containing haem group was identified as a somewhat larger electron-density feature. The structure determination of haemoglobin at 5.5 Å resolution in 1959 (Cullis *et al.*, 1962) showed that each of its two independent chains, α and β, had a fold similar to that of myoglobin and, thus, suggested a divergent evolutionary process for oxygen transport molecules. These first protein structures were mostly helical, features that could be recognized readily at low resolution. Had the first structures been of mostly β structure, as is the case for pepsin or chymotrypsin, history might have been different.

The absolute hand of the haemoglobin structure was determined using anomalous dispersion (Cullis *et al.*, 1962) in a manner similar to that used by Bijvoet. This was confirmed almost immediately when a 2 Å-resolution map of myoglobin was calculated in 1959 (Kendrew *et al.*, 1960). By plotting the electron density of the α-helices on cylindrical sections (Fig. 1.2.5.2), it was possible to see not only that the Pauling prediction of the α-helix structure was accurately obeyed, but also that the C_β atoms were consistent with *laevo* amino acids and that all eight helices were right-handed on account of the steric hindrance that would occur between the C_β atom and carbonyl oxygen in left-handed helices.

The first enzyme structure to be solved was that of lysozyme in 1965 (Blake *et al.*, 1965), following a gap of six years after the excitement caused by the discovery of the globin structures. Diffusion of substrates into crystals of lysozyme showed how substrates bound to the enzyme (Blake, Johnson *et al.*, 1967), which in turn suggested a catalytic mechanism and identified the essential catalytic residues.

From 1965 onwards, the rate of protein-structure determinations gradually increased to about one a year: carboxypeptidase (Reeke *et al.*, 1967), chymotrypsin (Matthews *et al.*, 1967), ribonuclease (Kartha *et al.*, 1967; Wyckoff *et al.*, 1967), papain (Drenth *et al.*, 1968), insulin (Adams *et al.*, 1969), lactate dehydrogenase (Adams *et al.*, 1970) and cytochrome *c* (Dickerson *et al.*, 1971) were early examples. Every new structure was a major event. These structures laid the foundation for structural biology. From a crystallographic point of view, Drenth's structure determination of papain was particularly significant in that he was able to show an amino-acid sequencing error where 13 residues had to be inserted between Phe28 and Arg31, and he showed that a 38-residue peptide that had been assigned to position 138 to 176 needed to be transposed to a position between the extra 13 residues and Arg31.

The structures of the globins had suggested that proteins with similar functions were likely to have evolved from a common precursor and, hence, that there might be a limited number of protein folding motifs. Comparison of the active centres of chymotrypsin and subtilisin showed that convergent evolutionary pathways could exist (Drenth *et al.*, 1972; Kraut *et al.*, 1972).

The variety of structures that were being studied increased rapidly. The first tRNA structures were determined in the 1960s (Kim *et al.*, 1973; Robertus *et al.*, 1974), the first spherical virus structure was published in 1978 (Harrison *et al.*, 1978) and the photoreaction centre membrane protein structure appeared in 1985 (Deisenhofer *et al.*, 1985). The rate of new structure determinations has continued to increase exponentially. In 1996, about one new structure was published every day. Partly in anticipation and partly to assure the availability of results, the Brookhaven Protein Data Bank (PDB) was brought to life at the 1971 Cold Spring Harbor Meeting (H. Berman & J. L. Sussman, private communication). Initially, it was difficult to persuade authors to submit their coordinates, but gradually this situation changed to one where most journals require coordinate submission to the PDB, resulting in today's access to structural results *via* the World Wide Web.

The growth of structural information permitted generalizations, such as that β-sheets have a left-handed twist when going from one strand to the next (Chothia, 1973) and that 'cross-over' β-α-β turns were almost invariably right-handed (Richardson, 1977). These observations and the growth of the PDB have opened up a new field of science. Among the many important results that have emerged from this wealth of data is a careful measurement of the main-chain dihedral angles, confirming the predictions of Ramachandran (Ramachandran & Sasisekharan, 1968), and of side-chain rotamers (Ponder & Richards, 1987). Furthermore, it is now possible to determine whether the folds of domains in a new structure relate to any previous results quite conveniently (Murzin *et al.*, 1995; Holm & Sander, 1997).

1.2.6. Technological developments (1958 to the 1980s)

In the early 1960s, there were very few who had experience in solving a protein structure. Thus, almost a decade passed after the success with the globins before there was a noticeable surge of new structure reports. In the meantime, there were persistent attempts to find alternative methods to determine protein structure.

Blow & Rossmann (1961) demonstrated the power of the single isomorphous replacement method. While previously it had been thought that it was necessary to have at least two heavy-atom compounds, if not many more, they showed that a good representation of the structure of haemoglobin could have been made by using only one good derivative. There were also early attempts at exploiting anomalous dispersion for phase determination. Rossmann (1961) showed that anomalous differences could be used to calculate a 'Bijvoet Patterson' from which the sites of the anomalous scatterers (and, hence, heavy-atom sites) could be determined. Blow & Rossmann (1961), North (1965) and Matthews (1966) used anomalous-dispersion data to help in phase determination. Hendrickson stimulated further interest by using Cu $K\alpha$ radiation and employing the anomalous effect of sulfur atoms in cysteines to solve the entire structure of the crambin molecule (Hendrickson & Teeter, 1981). With today's availability of synchrotrons, and hence the ability to tune to absorption edges, these early attempts to utilize anomalous data have been vastly extended to the powerful multiple-wavelength anomalous dispersion (MAD) method (Hendrickson, 1991). More recently, the generality of the MAD technique has been greatly expanded by using proteins in which methionine residues have been replaced by selenomethionine, thus introducing selenium atoms as anomalous scatterers.

Another advance was the introduction of the 'molecular replacement' technique (Rossmann, 1972). The inspiration for this method arose out of the observation that many larger proteins (*e.g.* haemoglobin) are oligomers of identical subunits and that many proteins can crystallize in numerous different forms. Rossmann & Blow (1962) recognized that an obvious application of the technique would be to viruses with their icosahedral symmetry. They pointed out that the symmetry of the biological oligomer can often be, and sometimes must be, 'noncrystallographic' or 'local', as opposed to being true for the whole infinite crystal lattice. Although the conservation of folds had become apparent in the study of the globins and a little later in the study of dehydrogenases (Rossmann *et al.*, 1974), in the 1960s the early development of the molecular replacement technique was aimed primarily at *ab initio* phase determination (Rossmann & Blow, 1963; Main & Rossmann, 1966; Crowther, 1969). It was only in the 1970s, when more structures became available, that it was possible to use the technique to solve homologous structures with suitable search models. Initially, there was a good deal of resistance to the use of the molecular replacement technique. Results from the rotation function were often treated with scepticism, the translation problem was thought to have no definitive answer, and there were excellent reasons to consider that phasing was impossible except for centric reflections (Rossmann, 1972). It took 25 years before the full power of all aspects of the molecular replacement technique was fully utilized and accepted (Rossmann *et al.*, 1985).

The first real success of the rotation function was in finding the rotational relationship between the two independent insulin monomers in the *P*3 unit cell (Dodson *et al.*, 1966). Crowther produced the fast rotation function, which reduced the computational times to manageable proportions (Crowther, 1972). Crowther (1969) and Main & Rossmann (1966) were able to formulate the problem of phasing in the presence of noncrystallographic symmetry in terms of a simple set of simultaneous complex equations. However, real advances came with applying the conditions of noncrystallographic symmetry in real space, which was the key to the solution of glyceraldehyde-3-phosphate dehydrogenase (Buehner *et al.*, 1974), tobacco mosaic virus disk protein (Bloomer *et al.*, 1978) and other structures, aided by Gerard Bricogne's program for electron-density averaging (Bricogne, 1976), which became a standard of excellence.

No account of the early history of protein crystallography is complete without a mention of ways of representing electron density. The 2 Å map of myoglobin was interpreted by building a model (on a scale of 5 cm to 1 Å with parts designed by Corey and Pauling at the California Institute of Technology) into a forest of vertical rods decorated with coloured clips at each grid point,

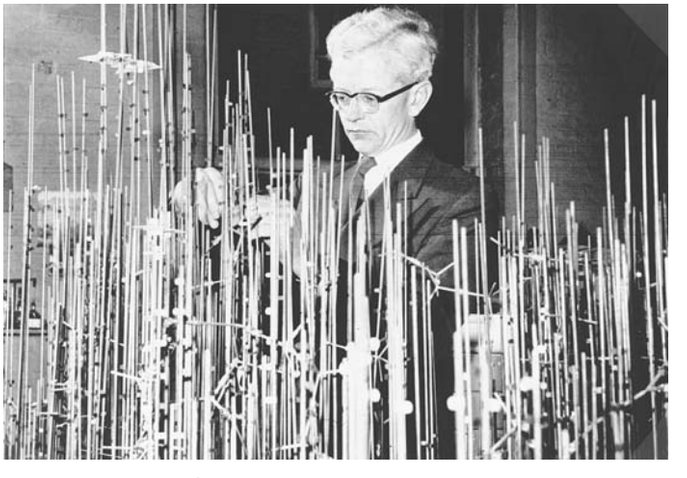

Fig. 1.2.6.1. The 2 Å-resolution map of sperm-whale myoglobin was represented by coloured Meccano-set clips on a forest of vertical rods. Each clip was at a grid point. The colour of the clip indicated the height of the electron density. The density was interpreted in terms of 'Corey–Pauling' models on a scale of 5 cm = 1 Å. Pictured is John Kendrew. (This figure was provided by M. F. Perutz.)

representing the height of the electron density (Fig. 1.2.6.1). Later structures, such as those of lysozyme and carboxypeptidase, were built with 'Kendrew' models (2 cm to 1 Å) based on electron-density maps displayed as stacks of large Plexiglas sheets. A major advance came with Fred Richards' invention of the optical comparator (a 'Richards box' or 'Fred's folly') in which the model was optically superimposed onto the electron density by reflection of the model in a half-silvered mirror (Richards, 1968). This allowed for convenient fitting of model parts and accurate placement of atoms within the electron density. The Richards box was the forerunner of today's computer graphics, originally referred to as an 'electronic Richards box'. The development of computer graphics for model building was initially met with reservation, but fortunately those involved in these developments persevered. Various programs became available for model building in a computer, but the undoubted champion of this technology was *FRODO*, written by Alwyn Jones (Jones, 1978).

1.2.7. Meetings

The birth of protein crystallography in the 1950s coincided with the start of the jet age, making attendance at international meetings far easier. Indeed, the number and variety of meetings have proliferated as much as the number and variety of structures determined. A critical first for protein crystallography was a meeting held at the Hirschegg ski resort in Austria in 1966. This was organized by Max Perutz (Cambridge) and Walter Hoppe (München). About 40 scientists from around the world attended, as well as a strong representation of students (including Robert Huber) from the Münich laboratory. The first Hirschegg meeting occurred just after the structure determination of lysozyme, which helped lift the cloud of pessimism after the long wait for a new structure since the structures of the globins had been solved in the 1950s. The meeting was very much a family affair where most attendees stayed an extra few days for additional skiing. The times were more relaxed in comparison with those of today's jet-setting scientists flying directly from synchrotron to international meeting, making ever more numerous important discoveries. A second Hirschegg meeting occurred two years later, but this time the number of participants had doubled. By 1970, the meeting had to be transferred to the village of Alpbach, which had more accommodation; however, most of the participants still knew each other.

Another set of international meetings (or schools, as the Italians preferred to call them) was initiated by the Italian crystallographers in 1976 at Erice, a medieval hilltop town in Sicily. These meetings have since been repeated every six years. The local organizer was Lodovico Riva di Sanseverino, whose vivacious sensitivity instilled a feeling of international fellowship into the rapidly growing number of structural biologists. The first meeting lasted two whole weeks, a length of time that would no longer be acceptable in today's hectic, competitive atmosphere.

It took time for the staid organizers of the IUCr triennial congress to recognize the significance of macromolecular structure. Thus, for many years, the IUCr meetings were poorly attended by structural biologists. However, recent meetings have changed, with biological topics representing about half of all activities. Nevertheless, the size of these meetings and their lack of focus have led to numerous large and small specialized meetings, from small Gordon Conferences and East and West Coast crystallography meetings in the USA to huge international congresses in virology, biochemistry and other sciences.

The publication of this volume by the International Union of Crystallography, the first volume of *International Tables* devoted to macromolecular crystallography, strongly attests to the increasing importance of this vital area of science.

Acknowledgements

I am gratefully indebted to Sharon Wilder, who has done much of the background checking that was required to write this chapter. Furthermore, she has been my permanent and faithful helper during a time that is, in part, covered in the review. I am additionally indebted to Max Perutz and David Davies, both of whom read the manuscript very carefully, making it possible to add a few personal accounts. I also thank the National Institutes of Health and the National Science Foundation for generous financial support.

phous replacement and anomalous-scattering measurements, are currently a mainstay, and will be for the foreseeable future. New isomorphous heavy-atom reagents and preparation methods will emerge; witness the valuable engineering of derivatives *via* mutagenesis to add such heavy-atom-binding sites as cysteine residues in proteins. Anomalous-scattering measurements from macromolecular crystals containing heavy metals or SeMet replacements of methionine residues in proteins have led to tremendous acceleration of the phase-problem solution for many structures – especially in the last two decades with the availability of 'tunable' synchrotron radiation. It should become possible to take better advantage of anomalous-scattering effects from lighter atoms already present in biological molecules: sulfur, phosphorus, oxygen, nitrogen and carbon. The increasingly higher intensity of synchrotron-radiation sources may permit structure solution from microcrystals of macromolecules. Incorporation of non-standard amino acids into proteins will become more common, leading to a vast array of new substitution possibilities. Molecular-replacement phasing from similar structures or from noncrystallographically redundant data is now commonplace and is continuing to become easier. Systematic molecular replacement using all known structures from databases may prove surprisingly powerful, if we can learn how to position small molecular fragments reliably. Systematic molecular-replacement approaches should help identify what folds may be present in a crystalline protein of unknown structure. Direct computational assaults on the phase problem are also becoming more aggressive and successful, although direct-methods approaches still work best for small macromolecular structures with very high resolution data.

Crystallography and structural biology have been helping to drive advances in three-dimensional visualization technology. Versatile molecular-graphics packages have been among the most important software applications for the best three-dimensional graphics workstations. Now that personal computers are being mass-produced with similar graphics capabilities, we can expect to see a molecular-graphics workstation at every computer, whether desktop or portable (terms that soon may become antiquated since everything will become more compact). Modes of input will include direct access to thought processes, and computer output devices will extend beyond light and sound. Universal internet access will provide immediate access to the rapidly increasing store of molecular information. As a result, we will achieve a more thorough understanding of patterns present in macromolecular structures: common folds of proteins and nucleic acids, three-dimensional motifs, and evolutionary relations among molecules. Simulations of complex molecular motions and interactions will be easier to display, making movies of molecules in motion commonplace. Facile 'virtual reality' representation of molecules will be a powerful research and teaching aid. Chemical reaction mechanisms will become better understood over time through interplay among theory, experiment and simulation. The ability to simulate all coupled chemical reactions in living cells and organisms will be achieved over time.

Advances in computational productivity depend on the intricate co-evolution of hardware and software. For silicon-based transistor chips, raw computational speed doubles approximately every 18 months (Moore's law). Tools and software for writing software will continue to advance rapidly. With greater modularity of software tools, it will become easier to coordinate existing programs and program suites. Enhanced automation, parallelization and development of new algorithms will also increase speed and throughput. More powerful software heuristics involving artificial intelligence, expert systems, neural nets and the like may permit unexpected advances in our understanding of the natural world.

In summary, we eagerly await what the future of science and of studies of molecular structure will bring. There is every reason to expect the unexpected. If the past is a guide, many new flowers will bloom to colour our world in bright new ways.

1.4.2. *Brief comments on Gazing into the crystal ball*

(M. G. ROSSMANN)

Edward Arnold and I had planned to write a joint commentary about our vision of the future of macromolecular crystallography. However, when Eddy produced the first draft of '*Perspectives for the future*', I was fascinated by his wide vision. I felt it more appropriate and far more interesting to make my own brief comments, stimulated by Eddy's observations.

When I was a graduate student in Scotland in the 1950s, physics departments were called departments of 'Natural Philosophy'. Clearly, the original hope had been that some aspects of science were all encompassing and gave insight to every aspect of observations of natural phenomena. However, in the twentieth century, with rapidly increasing technological advances, it appeared to be more and more difficult for any one person to study all of science. Disciplines were progressively subdivided. Learning became increasingly specialized. *International Tables* were created, and updated, for the use of a highly specialized and small community of crystallographic experts.

As I read Eddy's draft article, I became fascinated by the wide impact he envisioned for crystallography in the next few decades. Indeed, the lay person, reading his article, would barely be aware that this was an article anticipating the future impact of crystallography. The average reader would think that the topic was the total impact of science on our civilization. Thus, to my delight, I saw that crystallography might now be a catalyst for the reunification of fragmented science into a coherent whole. I therefore hope that these new *Tables* commissioned by the International Union of Crystallography may turn out to be a significant help to further the trend implied in Eddy's article.

References

1.2

Adams, M. J., Blundell, T. L., Dodson, E. J., Dodson, G. G., Vijayan, M., Baker, E. N., Harding, M. M., Hodgkin, D. C., Rimmer, B. & Sheat, S. (1969). *Structure of rhombohedral 2 zinc insulin crystals. Nature (London)*, **224**, 491–495.

Adams, M. J., Ford, G. C., Koekoek, R., Lentz, P. J. Jr, McPherson, A. Jr, Rossmann, M. G., Smiley, I. E., Schevitz, R. W. & Wonacott, A. J. (1970). *Structure of lactate dehydrogenase at 2.8 Å resolution. Nature (London)*, **227**, 1098–1103.

Astbury, W. T. (1933). *Fundamentals of fiber structure.* Oxford University Press.

Astbury, W. T., Mauguin, C., Hermann, C., Niggli, P., Brandenberger, E. & Lonsdale-Yardley, K. (1935). *Space groups*. In *Internationale Tabellen zur Bestimmung von Kristallstrukturen*, edited by C. Hermann, Vol. 1, pp. 84–87. Berlin: Begrüder Borntraeger.

Beevers, C. A. & Lipson, H. (1934). *A rapid method for the summation of a two-dimensional Fourier series. Philos. Mag.* **17**, 855–859.

Bernal, J. D. (1933). *Comments. Chem. Ind.* **52**, 288.

1.2 (*cont.*)

Bernal, J. D. & Crowfoot, D. (1934). *X-ray photographs of crystalline pepsin. Nature (London)*, **133**, 794–795.

Bernal, J. D. & Fankuchen, I. (1941). *X-ray and crystallographic studies of plant virus preparations. J. Gen. Physiol.* **25**, 111–165.

Bernal, J. D., Fankuchen, I. & Perutz, M. (1938). *An X-ray study of chymotrypsin and haemoglobin. Nature (London)*, **141**, 523–524.

Bijvoet, J. M. (1949). *K. Ned. Akad. Wet.* **52**, 313.

Bijvoet, J. M., Peerdeman, A. F. & van Bommel, A. J. (1951). *Determination of the absolute configuration of optically active compounds by means of X-rays. Nature (London)*, **168**, 271–272.

Blake, C. C. F., Johnson, L. N., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1967). *Crystallographic studies of the activity of hen egg-white lysozyme. Proc. R. Soc. London Ser. B*, **167**, 378–388.

Blake, C. C. F., Koenig, D. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1965). *Structure of hen egg-white lysozyme. A three-dimensional Fourier synthesis at 2 Å resolution. Nature (London)*, **206**, 757–761.

Blake, C. C. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1967). *On the conformation of the hen egg-white lysozyme molecule. Proc. R. Soc. London. Ser. B*, **167**, 365–377.

Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R. & Klug, A. (1978). *Protein disk of tobacco mosaic virus at 2.8 Å resolution showing the interactions within and between subunits. Nature (London)*, **276**, 362–368.

Blow, D. M. & Crick, F. H. C. (1959). *The treatment of errors in the isomorphous replacement method. Acta Cryst.* **12**, 794–802.

Blow, D. M. & Rossmann, M. G. (1961). *The single isomorphous replacement method. Acta Cryst.* **14**, 1195–1202.

Bluhm, M. M., Bodo, G., Dintzis, H. M. & Kendrew, J. C. (1958). *The crystal structure of myoglobin. IV. A Fourier projection of sperm-whale myoglobin by the method of isomorphous replacement. Proc. R. Soc. London Ser. A*, **246**, 369–389.

Bodo, G., Dintzis, H. M., Kendrew, J. C. & Wyckoff, H. W. (1959). *The crystal structure of myoglobin. V. A low-resolution three-dimensional Fourier synthesis of sperm-whale myoglobin crystals. Proc. R. Soc. London Ser. A*, **253**, 70–102.

Boyes-Watson, J., Davidson, E. & Perutz, M. F. (1947). *An X-ray study of horse methaemoglobin. I. Proc. R. Soc. London Ser. A*, **191**, 83–132.

Bragg, L. & Perutz, M. F. (1952). *The structure of haemoglobin. Proc. R. Soc. London Ser. A*, **213**, 425–435.

Bragg, W. H. (1915). *X-rays and crystal structure. Philos. Trans. R. Soc. London Ser. A*, **215**, 253–274.

Bragg, W. H. & Bragg, W. L. (1913). *The reflection of X-rays in crystals. Proc. R. Soc. London Ser. A*, **88**, 428–438.

Bragg, W. L. (1913). *The structure of some crystals as indicated by their diffraction of X-rays. Proc. R. Soc. London Ser. A*, **89**, 248–277.

Bragg, W. L. (1929*a*). *The determination of parameters in crystal structures by means of Fourier series. Proc. R. Soc. London Ser. A*, **123**, 537–559.

Bragg, W. L. (1929*b*). *An optical method of representing the results of X-ray analysis. Z. Kristallogr.* **70**, 475–492.

Bragg, W. L. (1958). *The determination of the coordinates of heavy atoms in protein crystals. Acta Cryst.* **11**, 70–75.

Bricogne, G. (1976). *Methods and programs for direct-space exploitation of geometric redundancies. Acta Cryst.* A**32**, 832–847.

Buehner, M., Ford, G. C., Moras, D., Olsen, K. W. & Rossmann, M. G. (1974). *Structure determination of crystalline lobster* D*-glyceraldehyde-3-phosphate dehydrogenase. J. Mol. Biol.* **82**, 563–585.

Chothia, C. (1973). *Conformation of twisted β-pleated sheets in proteins. J. Mol. Biol.* **75**, 295–302.

Cochran, W., Crick, F. H. C. & Vand, V. (1952). *The structure of synthetic polypeptides. I. The transform of atoms on a helix. Acta Cryst.* **5**, 581–586.

Crowfoot, D. (1948). *X-ray crystallographic studies of compounds of biochemical interest. Annu. Rev. Biochem.* **17**, 115–146.

Crowfoot, D., Bunn, C. W., Rogers-Low, B. W. & Turner-Jones, A. (1949). *The X-ray crystallographic investigation of the structure of penicillin.* In *The chemistry of penicillin*, edited by H. T. Clarke, J. R. Johnson & R. Robinson, pp. 310–367. Princeton University Press.

Crowther, R. A. (1969). *The use of non-crystallographic symmetry for phase determination. Acta Cryst.* B**25**, 2571–2580.

Crowther, R. A. (1972). *The fast rotation function.* In *The molecular replacement method*, edited by M. G. Rossmann, pp. 173–178. New York: Gordon & Breach.

Cullis, A. F., Muirhead, H., Perutz, M. F., Rossmann, M. G. & North, A. C. T. (1962). *The structure of haemoglobin. IX. A three-dimensional Fourier synthesis at 5.5 Å resolution: description of the structure. Proc. R. Soc. London Ser. A*, **265**, 161–187.

Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. (1985). *Structure of the protein subunits in the photosynthetic reaction centre of Rhodopseudomonas viridis at 3 Å resolution. Nature (London)*, **318**, 618–624.

Dickerson, R. E., Kendrew, J. C. & Strandberg, B. E. (1961). *The crystal structure of myoglobin: phase determination to a resolution of 2 Å by the method of isomorphous replacement. Acta Cryst.* **14**, 1188–1195.

Dickerson, R. E., Takano, T., Eisenberg, D., Kallai, O. B., Samson, L., Cooper, A. & Margoliash, E. (1971). *Ferricytochrome c. I. General features of the horse and bonito proteins at 2.8 Å resolution. J. Biol. Chem.* **246**, 1511–1535.

Dodson, E., Harding, M. M., Hodgkin, D. C. & Rossmann, M. G. (1966). *The crystal structure of insulin. III. Evidence for a 2-fold axis in rhombohedral zinc insulin. J. Mol. Biol.* **16**, 227–241.

Drenth, J., Hol, W. G. J., Jansonius, J. N. & Koekoek, R. (1972). *A comparison of the three-dimensional structures of subtilisin BPN′ and subtilisin Novo. Cold Spring Harbor Symp. Quant. Biol.* **36**, 107–116.

Drenth, J., Jansonius, J. N., Koekoek, R., Swen, H. M. & Wolthers, B. G. (1968). *Structure of papain. Nature (London)*, **218**, 929–932.

Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). *The structure of haemoglobin. IV. Sign determination by the isomorphous replacement method. Proc. R. Soc. London Ser. A*, **225**, 287–307.

Hahn, Th. (1983). Editor. *International tables for crystallography*, Vol. A. *Space-group symmetry.* Dordrecht: Kluwer Academic Publishers.

Harker, D. (1956). *The determination of the phases of the structure factors of non-centrosymmetric crystals by the method of double isomorphous replacement. Acta Cryst.* **9**, 1–9.

Harker, D. & Kasper, J. S. (1947). *Phases of Fourier coefficients directly from crystal diffraction data. J. Chem. Phys.* **15**, 882–884.

Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. & Bricogne, G. (1978). *Tomato bushy stunt virus at 2.9 Å resolution. Nature (London)*, **276**, 368–373.

Hendrickson, W. A. (1991). *Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. Science*, **254**, 51–58.

Hendrickson, W. A. & Teeter, M. M. (1981). *Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulphur. Nature (London)*, **290**, 107–113.

Henry, N. F. M. & Lonsdale, K. (1952). Editors. *International tables for X-ray crystallography*, Vol. I. *Symmetry groups.* Birmingham: Kynoch Press.

Hermann, C. (1935). Editor. *Internationale Tabellen zur Bestimmung von Kristallstrukturen*, 1. Band. *Gruppentheoretische Tafeln.* Berlin: Begrüder Borntraeger.

Hilton, H. (1903). *Mathematical crystallography and the theory of groups of movements.* Oxford: Clarendon Press.

Hodgkin, D. C., Kamper, J., Lindsey, J., MacKay, M., Pickworth, J., Robertson, J. H., Shoemaker, C. B., White, J. G., Prosen, R. J. & Trueblood, K. N. (1957). *The structure of vitamin B_{12}. I. An outline of the crystallographic investigation of vitamin B_{12}. Proc. R. Soc. London Ser. A*, **242**, 228–263.

1.2 (*cont.*)

Holm, L. & Sander, C. (1997). *New structure – novel fold? Structure*, **5**, 165–171.

Holmes, K. C., Stubbs, G. J., Mandelkow, E. & Gallwitz, U. (1975). *Structure of tobacco mosaic virus at 6.7 Å resolution. Nature (London)*, **254**, 192–196.

Hooke, R. (1665). *Micrographia*, p. 85. London: Royal Society.

Huygens, C. (1690). *Traité de la lumière.* Leiden. (English translation by S. P. Thompson, 1912. London: Macmillan and Co.)

Jones, T. A. (1978). *A graphics model building and refinement system for macromolecules. J. Appl. Cryst.* **11**, 268–272.

Judson, H. F. (1979). *The eighth day of creation.* New York: Simon & Schuster.

Kartha, G., Bello, J. & Harker, D. (1967). *Tertiary structure of ribonuclease. Nature (London)*, **213**, 862–865.

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. & Phillips, D. C. (1958). *A three-dimensional model of the myoglobin molecule obtained by X-ray analysis. Nature (London)*, **181**, 662–666.

Kendrew, J. C., Dickerson, R. E., Strandberg, B. E., Hart, R. G., Davies, D. R., Phillips, D. C. & Shore, V. C. (1960). *Structure of myoglobin. A three-dimensional Fourier synthesis at 2 Å resolution. Nature (London)*, **185**, 422–427.

Kim, S. H., Quigley, G. J., Suddath, F. L., McPherson, A., Sneden, D., Kim, J. J., Weinzierl, J. & Rich, A. (1973). *Three-dimensional structure of yeast phenylalanine transfer RNA: folding of the polynucleotide chain. Science*, **179**, 285–288.

Kraut, J., Robertus, J. D., Birktoft, J. J., Alden, R. A., Wilcox, P. E. & Powers, J. C. (1972). *The aromatic substrate binding site in subtilisin BPN′ and its resemblance to chymotrypsin. Cold Spring Harbor Symp. Quant. Biol.* **36**, 117–123.

Linderström-Lang, K. U. & Schellman, J. A. (1959). *Protein structure and enzyme activity.* In *The enzymes*, edited by P. D. Boyer, Vol. 1, 2nd ed., pp. 443–510. New York: Academic Press.

Lipson, H. & Cochran, W. (1953). *The determination of crystal structures*, pp. 325–326. London: G. Bell and Sons Ltd.

Lonsdale, K. (1928). *The structure of the benzene ring. Nature (London)*, **122**, 810.

Lonsdale, K. (1936). *Structure factor tables.* London: Bell.

McLachlan, D. Jr & Champaygne, E. F. (1946). *A machine for the application of sand in making Fourier projections of crystal structures. J. Appl. Phys.* **17**, 1006–1014.

Main, P. & Rossmann, M. G. (1966). *Relationships among structure factors due to identical molecules in different crystallographic environments. Acta Cryst.* **21**, 67–72.

Matthews, B. W. (1966). *The determination of the position of anomalously scattering heavy atom groups in protein crystals. Acta Cryst.* **20**, 230–239.

Matthews, B. W., Sigler, P. B., Henderson, R. & Blow, D. M. (1967). *Three-dimensional structure of tosyl-α-chymotrypsin. Nature (London)*, **214**, 652–656.

Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). *SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol.* **247**, 536–540.

North, A. C. T. (1965). *The combination of isomorphous replacement and anomalous scattering data in phase determination of non-centrosymmetric reflexions. Acta Cryst.* **18**, 212–216.

Olby, R. (1974). *The path to the double helix.* Seattle: University of Washington Press.

Patterson, A. L. (1934). *A Fourier series method for the determination of the components of interatomic distances in crystals. Phys. Rev.* **46**, 372–376.

Patterson, A. L. (1935). *A direct method for the determination of the components of interatomic distances in crystals. Z. Kristallogr.* **90**, 517–542.

Pauling, L. & Corey, R. B. (1951). *Configurations of polypeptide chains with favored orientations around single bonds: two new pleated sheets. Proc. Natl Acad. Sci. USA*, **37**, 729–740.

Pauling, L., Corey, R. B. & Branson, H. R. (1951). *The structure of proteins: two hydrogen-bonded helical configurations of the polypeptide chain. Proc. Natl Acad. Sci. USA*, **37**, 205–211.

Pepinsky, R. (1947). *An electronic computer for X-ray crystal structure analysis. J. Appl. Phys.* **18**, 601–604.

Perutz, M. F. (1946). *Trans. Faraday Soc.* **42**B, 187.

Perutz, M. F. (1949). *An X-ray study of horse methaemoglobin. II. Proc. R. Soc. London Ser. A*, **195**, 474–499.

Perutz, M. F. (1951*a*). *New X-ray evidence on the configuration of polypeptide chains. Nature (London)*, **167**, 1053–1054.

Perutz, M. F. (1951*b*). *The 1.5-A. reflextion from proteins and polypeptides. Nature (London)*, **168**, 653–654.

Perutz, M. F. (1954). *The structure of haemoglobin. III. Direct determination of the molecular transform. Proc. R. Soc. London Ser. A*, **225**, 264–286.

Perutz, M. F. (1956). *Isomorphous replacement and phase determination in non-centrosymmetric space groups. Acta Cryst.* **9**, 867–873.

Perutz, M. F. (1962). *Proteins and nucleic acids. Structure and function.* Amsterdam: Elsevier Publishing Co.

Perutz, M. F. (1997). *Keilin and the molteno.* In *Selected topics in the history of biochemistry: personal reflections, V*, edited by G. Semenza & R. Jaenicke, pp. 57–67. Amsterdam: Elsevier.

Ponder, J. W. & Richards, F. M. (1987). *Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes. J. Mol. Biol.* **193**, 775–791.

Ramachandran, G. N. & Sasisekharan, V. (1968). *Conformation of polypeptides and proteins. Adv. Protein Chem.* **23**, 283–438.

Reeke, G. N., Hartsuck, J. A., Ludwig, M. L., Quiocho, F. A., Steitz, T. A. & Lipscomb, W. N. (1967). *The structure of carboxypeptidase A, VI. Some results at 2.0-Å resolution, and the complex with glycyl-tyrosine at 2.8-Å resolution. Proc. Natl Acad. Sci. USA*, **58**, 2220–2226.

Richards, F. M. (1968). *The matching of physical models to three-dimensional electron-density maps: a simple optical device. J. Mol. Biol.* **37**, 225–230.

Richardson, J. S. (1977). *β-Sheet topology and the relatedness of proteins. Nature (London)*, **268**, 495–500.

Robertson, J. M. (1935). *An X-ray study of the structure of phthalocyanines. Part I. The metal-free, nickel, copper, and platinum compounds. J. Chem. Soc.* pp. 615–621.

Robertson, J. M. (1936). *Numerical and mechanical methods in double Fourier synthesis. Philos. Mag.* **21**, 176–187.

Robertus, J. D., Ladner, J. E., Finch, J. T., Rhodes, D., Brown, R. S., Clark, B. F. C. & Klug, A. (1974). *Structure of yeast phenylalanine tRNA at 3 Å resolution. Nature (London)*, **250**, 546–551.

Rossmann, M. G. (1960). *The accurate determination of the position and shape of heavy-atom replacement groups in proteins. Acta Cryst.* **13**, 221–226.

Rossmann, M. G. (1961). *The position of anomalous scatterers in protein crystals. Acta Cryst.* **14**, 383–388.

Rossmann, M. G. (1972). Editor. *The molecular replacement method.* New York: Gordon & Breach.

Rossmann, M. G., Arnold, E., Erickson, J. W., Frankenberger, E. A., Griffith, J. P., Hecht, H. J., Johnson, J. E., Kamer, G., Luo, M., Mosser, A. G., Rueckert, R. R., Sherry, B. & Vriend, G. (1985). *Structure of a human common cold virus and functional relationship to other picornaviruses. Nature (London)*, **317**, 145–153.

Rossmann, M. G. & Blow, D. M. (1962). *The detection of sub-units within the crystallographic asymmetric unit. Acta Cryst.* **15**, 24–31.

Rossmann, M. G. & Blow, D. M. (1963). *Determination of phases by the conditions of non-crystallographic symmetry. Acta Cryst.* **16**, 39–45.

Rossmann, M. G., Moras, D. & Olsen, K. W. (1974). *Chemical and biological evolution of a nucleotide-binding protein. Nature (London)*, **250**, 194–199.

Sanger, F. & Thompson, E. O. P. (1953*a*). *The amino-acid sequence in the glycyl chain of insulin. 1. The identification of lower peptides from partial hydrolysates. Biochem. J.* **53**, 353–366.

1.2 (*cont.*)

Sanger, F. & Thompson, E. O. P. (1953*b*). *The amino-acid sequence in the glycyl chain of insulin. 2. The investigation of peptides from enzymic hydrolysates. Biochem. J.* **53**, 366–374.

Sanger, F. & Tuppy, H. (1951). *The amino-acid sequence in the phenylalanyl chain of insulin. 2. The investigation of peptides from enzymic hydrolysates. Biochem. J.* **49**, 481–490.

Schoenflies, A. M. (1891). *Krystallsysteme und Krystallstruktur.* Leipzig: Druck und Verlag von B. G. Teubner.

Snow, C. P. (1934). *The search.* Indianapolis and New York: Bobbs-Merrill Co.

Watson, J. D. (1954). *The structure of tobacco mosaic virus. I. X-ray evidence of a helical arrangement of sub-units around the longitudinal axis. Biochim. Biophys. Acta*, **13**, 10–19.

Watson, J. D. (1968). *The double helix.* New York: Atheneum.

Wyckoff, H. W., Hardman, K. D., Allewell, N. M., Inagami, T., Johnson, L. N. & Richards, F. M. (1967). *The structure of ribonuclease-S at 3.5 Å resolution. J. Biol. Chem.* **242**, 3984–3988.

1.3

Achari, A., Somers, D. O., Champness, J. N., Bryant, P. K., Rosemond, J. & Stammers, D. K. (1997). *Crystal structure of the anti-bacterial sulfonamide drug target dihydropteroate synthase. Nature Struct. Biol.* **4**, 490–497.

Adman, E. T., Stenkamp, R. E., Sieker, L. C. & Jensen, L. H. (1978). *A crystallographic model for azurin at 3 Å resolution. J. Mol. Biol.* **123**, 35–47.

Aertgeerts, K., De Bondt, H. L., De Ranter, C. J. & Declerck, P. J. (1995). *Mechanisms contributing to the conformational and functional flexibility of plasminogen activator inhibitor-1. Nature Struct. Biol.* **2**, 891–897.

Ævarsson, A., Chuang, J. L., Wynn, R. M., Turley, S., Chuang, D. T. & Hol, W. G. J. (2000). *Crystal structure of human branched-chain α-ketoacid dehydrogenase and the molecular basis of multienzyme complex deficiency in maple syrup urine disease. Struct. Fold. Des.* **8**, 277–291.

Allaire, M., Chernaia, M. M., Malcolm, B. A. & James, M. N. (1994). *Picornaviral 3C cysteine proteinases have a fold similar to chymotrypsin-like serine proteinases. Nature (London)*, **369**, 72–76.

Allured, V. S., Collier, R. J., Carroll, S. F. & McKay, D. B. (1986). *Structure of exotoxin A of Pseudomonas aeruginosa at 2.0-Å resolution. Proc. Natl Acad. Sci. USA*, **83**, 1320–1324.

Almassy, R. J. & Dickerson, R. E. (1978). *Pseudomonas cytochrome c551 at 2.0 Å resolution: enlargement of the cytochrome c family. Proc. Natl Acad. Sci. USA*, **75**, 2674–2678.

Amos, L. A. & Lowe, J. (1999). *How Taxol stabilises microtubule structure. Chem. Biol.* **6**, R65–R69.

Arnold, E., Das, K., Ding, J., Yadav, P. N., Hsiou, Y., Boyer, P. L. & Hughes, S. H. (1996). *Targeting HIV reverse transcriptase for anti-AIDS drug design: structural and biological considerations for chemotherapeutic strategies. Drug Des. Discov.* **13**, 29–47.

Arnold, E. & Rossmann, M. G. (1988). *The use of molecular-replacement phases for the refinement of the human rhinovirus 14 structure. Acta Cryst.* A**44**, 270–283.

Arnold, E. & Rossmann, M. G. (1990). *Analysis of the structure of a common cold virus, human rhinovirus 14, refined at a resolution of 3.0 Å. J. Mol. Biol.* **211**, 763–801.

Arnold, G. F. & Arnold, E. (1999). *Using combinatorial libraries to develop vaccines. ASM News*, **65**, 603–610.

Arnold, G. F., Resnick, D. A., Li, Y., Zhang, A., Smith, A. D., Geisler, S. C., Jacobo-Molina, A., Lee, W., Webster, R. G. & Arnold, E. (1994). *Design and construction of rhinovirus chimeras incorporating immunogens from polio, influenza, and human immunodeficiency viruses. Virology*, **198**, 703–708.

Arnoux, P., Haser, R., Izadi, N., Lecroisey, A., Delepierre, M., Wandersman, C. & Czjzek, M. (1999). *The crystal structure of HasA, a hemophore secreted by Serratia marcescens. Nature Struct. Biol.* **6**, 516–520.

Athanasiadis, A., Vlassi, M., Kotsifaki, D., Tucker, P. A., Wilson, K. S. & Kokkinidis, M. (1994). *Crystal structure of PvuII endonuclease reveals extensive structural homologies to EcoRV. Nature Struct. Biol.* **1**, 469–475.

Baca, A. M., Sirawaraporn, R., Turley, S., Athappilly, F., Sirawaraporn, W. & Hol, W. G. J. (2000). *Crystal structure of Mycobacterium tuberculosis 6-hydroxymethyl-1,8-dihydropteroate synthase in complex with pterin monophosphate: new insight into the enzymatic mechanism and sulfa-drug action. J. Mol. Biol.* **302**, 1193–1212.

Baldwin, E. T., Bhat, T. N., Gulnik, S., Hosur, M. V., Sowder, R. C. I., Cachau, R. E., Collins, J., Silva, A. M. & Erickson, J. W. (1993). *Crystal structures of native and inhibited forms of human cathepsin D: implications for lysosomal targeting and drug design. Proc. Natl Acad. Sci. USA*, **90**, 6796–6800.

Baldwin, E. T., Weber, I. T., St Charles, R., Xuan, J. C., Appella, E., Yamada, M., Matsushima, K., Edwards, B. F., Clore, G. M. & Gronenborn, A. M. (1991). *Crystal structure of interleukin 8: symbiosis of NMR and crystallography. Proc. Natl Acad. Sci. USA*, **88**, 502–506.

Baldwin, J. J., Ponticello, G. S., Anderson, P. S., Christy, M. E., Murcko, M. A., Randall, W. C., Schwam, H., Sugrue, M. F., Springer, J. P., Gautheron, P., Grove, J., Mallorga, P., Viader, M. P., McKeever, B. M. & Navia, M. A. (1989). *Thienothiopyran-2-sulfonamides: novel topically active carbonic anhydrase inhibitors for the treatment of glaucoma. J. Med. Chem.* **32**, 2510–2513.

Ban, N., Nissen, P., Hansen, J., Capel, M., Moore, P. B. & Steitz, T. A. (1999). *Placement of protein and RNA structures into a 5 Å-resolution map of the 50S ribosomal subunit. Nature (London)*, **400**, 841–847.

Banbula, A., Potempa, J., Travis, J., Fernandez-Catalan, C., Mann, K., Huber, R., Bode, W. & Medrano, F. (1998). *Amino-acid sequence and three-dimensional structure of the Staphylococcus aureus metalloproteinase at 1.72 Å resolution. Structure*, **6**, 1185–1193.

Banner, D. W., D'Arcy, A., Chene, C., Winkler, F. K., Guha, A., Konigsberg, W. H., Nemreson, Y. & Kirchhofer, D. (1996). *The crystal structure of the complex of blood coagulation factor VIIa with soluble tissue factor. Nature*, **380**, 41–46.

Banner, D. W., D'Arcy, A., Janes, W., Gentz, R., Schoenfeld, H. J., Broger, C., Loetscher, H. & Lesslauer, W. (1993). *Crystal structure of the soluble human 55 kd TNF receptor–human TNF beta complex: implications for TNF receptor activation. Cell*, **7**, 431–445.

Baumann, U. (1994). *Crystal structure of the 50 kDa metallo protease from Serratia marcescens. J. Mol. Biol.* **242**, 244–251.

Beaman, T. W., Binder, D. A., Blanchard, J. S. & Roderick, S. L. (1997). *Three-dimensional structure of tetrahydrodipicolinate N-succinyltransferase. Biochemistry*, **36**, 489–494.

Beaman, T. W., Sugantino, M. & Roderick, S. L. (1998). *Structure of the hexapeptide xenobiotic acetyltransferase from Pseudomonas aeruginosa. Biochemistry*, **37**, 6689–6696.

Becker, J. W., Marcy, A. I., Rokosz, L. L., Axel, M. G., Burbaum, J. J., Fitzgerald, P. M., Cameron, P. M., Esser, C. K., Hagmann, W. K., Hermes, J. D. & Springer, J. P. (1995). *Stromelysin-1: three dimensional structure of the inhibited catalytic domain and of the C-truncated proenzyme. Protein Sci.* **4**, 1966–1976.

Bentley, G., Dodson, E., Dodson, G., Hodgkin, D. & Mercola, D. (1976). *Structure of insulin in 4-zinc insulin. Nature (London)*, **261**, 166–168.

Bernstein, B. E., Williams, D. M., Bressi, J. C., Kuhn, P., Gelb, M. H., Blackburn, G. M. & Hol, W. G. J. (1998). *A bisubstrate analog induces unexpected conformational changes in phosphoglycerate kinase from Trypanosoma brucei. J. Mol. Biol.* **279**, 1137–1148.

Bernstein, F. C., Koetzle, T. F., Williams, G. J., Meyer, E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *The Protein Data Bank. A computer-based archival file for macromolecular structures. Eur. J. Biochem.* **80**, 319–324.

Betz, M., Huxley, P., Davies, S. J., Mushtaq, Y., Pieper, M., Tschesche, H., Bode, W. & Gomis-Ruth, F. X. (1997). *1.8-Å*