From the Structures of Simple Salts to Those of Sophisticated Viruses

Michael G. Rossmann

Department of Biological Sciences, Purdue University, West Lafayette, IN 47907-1392, USA.
E-mail: mgr@indiana.bio.purdue.edu

Abstract

An account is given of the development of X-ray crystallography from its beginnings in 1912 to its present successes in the determination of virus structures. Brief mention is made of early inorganic and subsequent organic molecule structure determinations, which eventually laid the foundations for the first protein structures in 1957–1959. Various experimental and computational advances were required for these successes. However, additional technical advances were necessary before the first near atomic resolution structures of spherical viruses became possible in 1978–1979. Current emphasis is in probing the multifunctional biological properties of viral capsids when they infect cells.

1. Introduction

No history has a beginning. There are always earlier events which stimulate what happens subsequently. Undoubtedly, virus crystallography had its origins in protein crystallography, which in turn grew from the study of organic molecules and, earlier still, of inorganic compounds. Nor is there a strict linear evolution of one topic to another, but there is extensive cross-fertilization of ideas, methods and results. As much as some would wish that scientific discovery is a rational progression of consecutive events, science is largely determined by personalities. For instance, it is probably not an accident that X-ray diffraction was first demonstrated in the town which had become home to Röntgen, von Laue and Ewald. But it was Bragg in Cambridge, with his physical mind and love of simplicity, who had the greatest impact on the following generation of crystallographers.

Michael Rossmann was born in Germany, but educated in England and the US. His graduate work was with John Monteath Robertson in Glasgow. His first postdoctoral experience was in the laboratory of Bill Lipscomb (the Colonel) after which he moved to Cambridge to work in Max Perutz's laboratory on the 5.5 Å resolution structure determination of horse hemoglobin. In 1964, he moved to Purdue University where he first worked on the structure of dehydrogenases and later on various viruses.

Here, I give my view of the more important events that led to our present extensive knowledge of virus structure. However, as in all histories, no two people are likely to tell the same story, particularly when the story tellers are recounting some of their own experiences [compare, for instance, the article by Perutz in *Methods in Enzymology* (Perutz, 1985)]. I have chosen to start – roughly – with the 1912 discovery in Munich of X-ray diffraction by crystals, but I hurry on to more recent events in order to expand a little on the successes of the last two decades. While I hope that this account can give some of the flavor of the interaction of people and ideas, it is a task beyond my available time and ability to balance accurately the factors which led to each new advance. I hope that the many actors who feature in this account will forgive me for the bias which I will surely have displayed in this hastily written text.

2. Pre-1912

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 can 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 the 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*.

3. 1912 to the 1950's

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 of 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 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 the *International Tables*, Volume 1 (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 1920's. 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 1930's. 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 Ni 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 previously 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, it was now determined what was the absolute structure of naturally occurring amino acids and riboses.

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 eye 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), made 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 in the 1950s (Hodgkin *et al.*, 1957). They first used a cobalt atom as a heavy atom on a vitamin B_{12} fragment and were able to recognize the 'corin' ring structure. This success was made possible by the extraordinary collaboration between Dorothy Hodgkin in Oxford, England, and Kenneth Trueblood at UCLA in Los Angeles, California, USA. 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 analyze the periodic electronic 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 Fourier analysis demonstrated that the central problem in structural crystallography was the phase problem. 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 analog machine of Ray Pepinsky (Pepinsky, 1947) and mechanical analog 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 labor for decades in 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 usual method of measuring intensities was by manual inspection of reflection 'spots' on films by comparison 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.

4. The first investigation of biological macromolecules

Leeds, in the county of Yorkshire, was one of the centers of the English fabric industry and was the home of a small research institute established to investigate the properties of natural fibers. 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 to crystallize the first enzyme, namely urease, in 1926. Not much later, J. D. Bernal and his student, Dorothy Crowfoot (Hodgkin), in Cambridge investigated crystals of pepsin. The resultant 1934 paper in *Nature (London)*, (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 that 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, 1953a,b).

Soon after Bernal and Hodgkin photographed an X-ray diffraction pattern of pepsin, Max Perutz started his historic investigation of hemoglobin.† 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 which suggested that hemoglobin 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.

Pauling in Pasadena (Pauling & Corey, 1951; Pauling *et al.*, 1951) was building helical polypeptide structures to explain Astbury's α patterns and perhaps to understand the helical structures in globular proteins such as hemoglobin. 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, 1951a,b). Demonstration of this reflection in horse hair, then in fibers of polybenzyl-L-glutamate, in muscle (with Hugh Huxley) and finally in hemoglobin crystals showed that Pauling's proposed α-helix really existed in hemoglobin and presumably also in other globular proteins. Confirmation of helical-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 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 by Bernal *et al.* (1938). I did the experimental work, (and) Bernal showed me how to interpret the X-ray pictures.'

Perutz's studies of helices in polybenzyl-L-glutamate and their presence in hemoglobin, the impact on the structure determination of nucleic acids was even more significant. The events leading to the discovery 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 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 the viral RNA protected by an equally helical array of proteins. Eventually, after continuing studies by Aaron Klug, Rosalind Franklin, Ken Holmes and others, the structure was determined at atomic resolution (Holmes *et al.*, 1975).

5. Globular proteins in the 1950s

Max Perutz had joined Sir Lawrence Bragg in 1936 in Cambridge. Inspired in part by Keilin (Perutz, 1997), Perutz started to study crystalline hemoglobin. This work was interrupted by World War II but, once the war had finished, Perutz tenaciously developed a series of highly ingenious techniques. All of these procedures have their counterparts in modern 'macromolecular crystallography', although today few 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 hemoglobin after drying them in air for short periods of time and then sealing them in capillaries. He found that there were at least seven consecutive shrinkage stages of the unit cell. He realised 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). Thus, he was able to determine the phases (signs) of the central part of the (*h*0*l*) reciprocal lattice. This technique is identical to the use of diffraction data from different unit cells for averaging electron density in the 'modern' molecular replacement method. In the hemoglobin 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 hemoglobin 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 skeptics 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 hemoglobin 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 non-centric three-dimensional data. It took another 5 years to determine the first globular protein to near atomic resolution.

---

† Perutz writes, 'I measured the absolute intensity of reflections 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.'

In 1950, David Harker was awarded one million dollars by the Damon Runyon Foundation 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 and, therefore, solve the structure by 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 in 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.

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 by Dickerson *et al.* (1961).

One problem of 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 hemoglobin. Perutz (1956) and Bragg (1958) had each made proposals of how to solve 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 hemoglobin by using a $(F_{H1} - F_{H2})^2$ 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.

6. The first protein structures (1957–1970s)

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. 2) in 1957 (Kendrew *et al.*, 1958). It clearly showed sausage-like features which were assumed to be α-helices. The Fe-containing heme group was identified as a somewhat larger electron density feature. The structure determination of hemoglobin in 1959 at 5.5 Å resolution (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
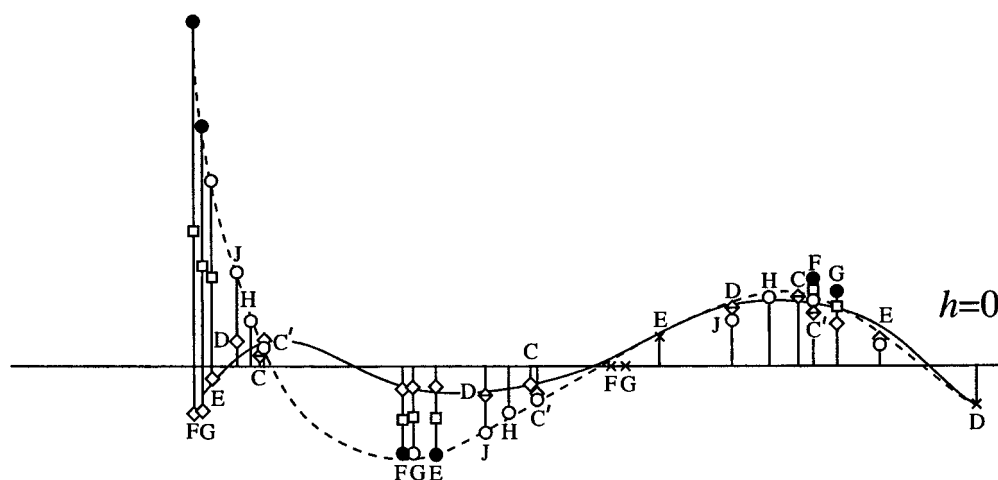
Fig. 1. Change of structure amplitude as a function of salt concentration of the suspension medium of the low-order $h0l$ reflections at various lattice shrinkage stages (C, C′, D, E, F, G, H, J). Reprinted with permission from Perutz (1954). Copyright (1954) The Royal Society of London.

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 hemoglobin structure was determined using anomalous dispersion (Cullis *et al.*, 1962) in a manner similar to that which had been 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. 3), 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), a gap of 6 years after the excitement caused by the discovery of the globin structures. Diffusion of substrate into crystals of lysozyme showed how substrate bound to the enzyme (Blake, Johnson *et al.*, 1967), which in turn suggested a catalytic mechanism and identified the essential catalytic residues. The use of structural arguments to support chemical observations was a major step.

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 of 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 that a 38-residue peptide, which had been assigned to position 138 to 176, needed to be transposed to a position between the extra 13 residues and Arg31.

The variety of structures that were being studied increased rapidly. The first tRNA structures were determined in the 1970s (Kim *et al.*, 1973; Robertus *et al.*, 1974), the first spherical virus structure was published in 1978 (Harrison *et al.*, 1978) and the photoreaction center membrane protein 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 ensure the availability of results, the Brookhaven Protein Data Bank (PDB) was brought to life at the 1971 Cold Spring Harbor Meeting (Berman & Sussman, 1998). Initially, it was difficult to persuade authors to submit their coordinates, but gradually this changed to where most journals require coordinate submission to the PDB, resulting in today's
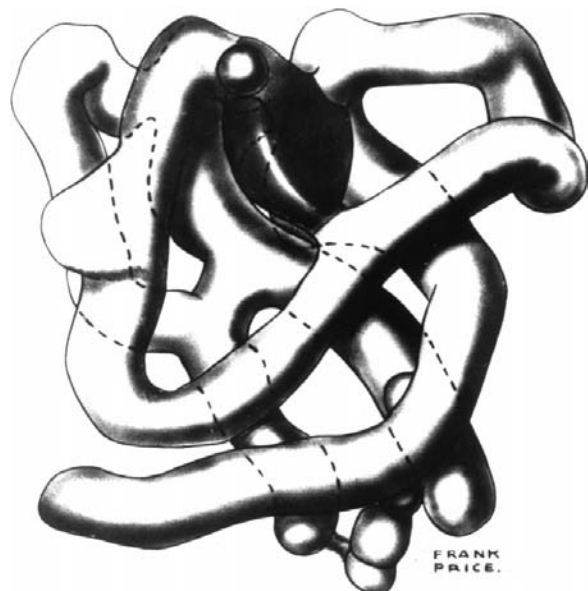

Fig. 2. A model of the myoglobin molecule at 6 Å resolution. Reprinted with permission from Bluhm *et al.* (1958). Copyright (1958) The Royal Society of London.
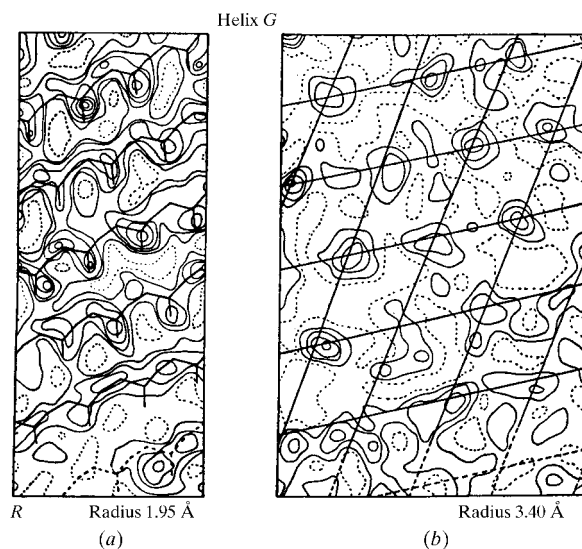

Fig. 3. Cylindrical sections through a helical segment of polypeptide chain. (*a*) The density in a cylinder mantle of 1.95 Å radius, corresponding to the mean radius of the main-chain atoms in an α-helix. The calculated atomic positions of the α-helix are superimposed and are seen to correspond roughly with 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 (J. C. Kendrew & J. D. Watson, unpublished results). Reprinted with permission from Perutz (1962). Copyright (1962) Elsevier Publishing Co.

World Wide Web accessibility to structural results. Among the many important results that have emerged from this wealth of data is a careful mapping 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 (Murzin *et al.*, 1995; Holm & Sander, 1997).

7. Technological developments (1958–1980s)

In the early 1960s, there were very few who had 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 anxious 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 hemoglobin 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 site 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).

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.* hemoglobin) 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, 'non-crystallographic' 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 dehy-

drogenases (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 skepticism, 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 $R3$ unit cell (Dodson *et al.*, 1966). This showed that there was a dimer in the asymmetric unit and that the unit cell probably contained a hexamer. Crowther produced the fast rotation function, which reduced the computational times to manageable proportions (Crowther, 1972). Crowther (1969) and also Main & Rossmann (1966) were able to formulate the problem of phasing in the presence of non-crystallographic symmetry in terms of a simple set of simultaneous complex equations. However, real advances came with applying the conditions of non-crystallographic 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. Electron density averaging was the first of numerous methods that employed density modification justified by physical expectations of the structure. One important example is solvent flattening, implicit in the regions of the cell not governed by non-crystallographic symmetry, but also applicable in the space between molecules in all structures (Wang, 1985).

No account of the early history of macromolecular 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 by colored clips at each grid point representing the height of the electron density (Fig. 4). 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's 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 easy 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 winner in this technology was *FRODO*, written by Alwyn Jones (Jones, 1978).

8. Viruses

As mentioned above, the first virus structure to be investigated was that of TMV in which the single strand of genomic RNA makes a helix on which are assembled $16\frac{2}{3}$ protective protein subunits per turn (Fig. 5). However, most viruses have a greater separation of protein and nucleic acid where the protein forms a spherical shell within which resides a single copy of the nucleic acid genome. Prior to 1953, electron microscope studies had shown this general organization, but little more. Watson & Crick (1953a) had only recently discovered the structure of DNA and suggested its biological significance (Watson & Crick, 1953b). The size of the codon was still completely unknown, but was recognized as requiring a minimum of two, and probably no more than six, bases. Then, Watson and Crick (Watson & Crick, 1953a) made the remarkable but simple observation that the viral nucleic acid would most probably be able to code only a few proteins of limited molecular weight. Each protein being identified would have to have the same environment and, hence, simple spherical viruses were likely to be regular polyhedra.

It had been known since the 1940s that some viruses could be crystallized and then redissolved to produce a solution containing infectious particles. For instance, Crowfoot & Schmidt (1945) had crystallized satellite tobacco necrosis virus (STNV) and shown that these crystals produced excellent X-ray diffraction patterns. Southern bean mosaic virus (SBMV) had also been crystallized by Price (1946). Caspar (1956) used single crystals of tomato bushy stunt virus (TBSV) to show, in another paper of great insight, that the plant tomato bushy stunt virus was, indeed, icosahedral, as had been predicted by Crick and Watson. But extensive electron microscopy and molecular weight data were starting to show that many spherical viruses had many more than 60 capsid protein subunits – 60 being the number of asymmetric units in an icosahedron, the largest of all the regular polyhedra. Thus, in 1962, Caspar & Klug (1962) published yet another remarkable paper in which they worked out rules of quasi-symmetry, which showed that 60 *T* subunits could make quasi-identical interactions to form the viral capsid. *T* was called the triangulation

Fig. 4. John Kendrew building the model of myoglobin. The 2 Å resolution map of sperm whale myoglobin was represented by colored mecano set clips on a forest of vertical rods. Each clip was at a grid point. The color of the clip indicated the height of electron density. The density was interpreted in terms of 'Corey–Pauling' models on a scale of 5 cm $\equiv$ 1 Å. (This figure was provided by M. F. Perutz.)

number, whose admissible values were strictly limited. These rules were found to be of great help in the interpretation of ever better electron-microscopy studies, although later higher-resolution X-ray studies have shown numerous larger and smaller exceptions. Nevertheless, the T number remains an essential first in the study of any virus structure.

Until the early 1970s, there was no real hope of ever determining an atomic resolution virus structure. Furthermore, most crystallographers were ignorant of the multifunctionally important role played by structure in a viral life cycle. But gradually technology improved to place a virus structure as a distant hope. Some of the important contributions were: development of oscillation photography, spearheaded by Uli Arndt and Alan Wonacott (Arndt & Wonacott, 1977); commercially available high-intensity rotating-anode X-ray generators; X-ray focusing mirrors (Franks, 1955; Harrison, 1968); computer controlled film scanning and digitizing instruments to cope with the large number of reflections created by the large unit cells; recognition that the presence of non-crystallographic symmetry would be a key to the phasing problem (Rossmann & Blow, 1962); development of the rotation function (Rossmann & Blow, 1962) for finding the orientation of virus particles in the cell; experience with real-space averaging for phase improvement (Buehner et al., 1974; Bricogne, 1976; Johnson, 1978); computers powerful enough to handle the automatic data processing, rotation function and electron density averaging procedures; and confidence in the power of protein crystallography.
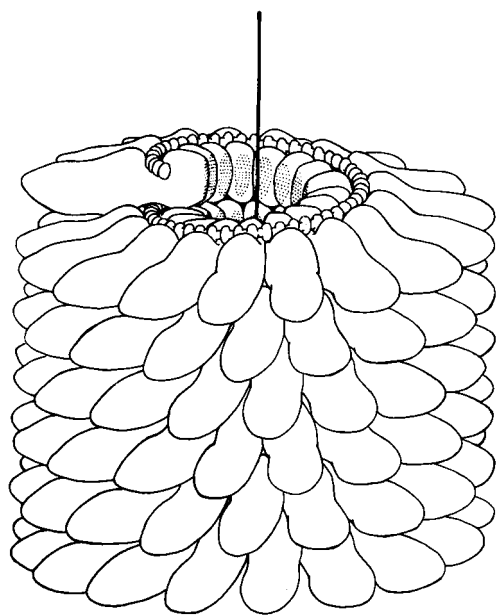

Fig. 5. Drawing of the structure of tobacco mosaic virus by Don Caspar. The helical chain represents RNA and the wooden shoes represent protein molecules. Reprinted with permission from Caspar (1963). Copyright (1963) Academic Press, Inc.

These techniques eventually led to the structure of the TMV disk (Bloomer et al., 1978) (a duplex assembly of 17 protein subunits) and its relevance to the fascinating assembly of TMV (Butler & Klug, 1978). In 1978, Harrison and his colleagues were able to obtain the first structure of an icosahedral virus, TBSV (Harrison et al., 1978). Its major interest was how the $T = 3$ symmetry could be accommodated. Shortly afterwards, we were able to obtain the structure of the $T = 3$ SBMV plant virus, which showed that the shell domain of TBSV was remarkably similar in structure to that of SBMV. Soon, other virus or viral capsid structures came along, such as STNV (Liljas et al., 1982) and the hexon of adenovirus (Burnett et al., 1985), all of which had one or two domains of a similar antiparallel β-barrel structure in their capsid proteins. Probably, these different viruses evolved from a common ancestor. Exceptions came in the form of the bacterial phage MS2 (Valegård et al., 1990) and the alphavirus capsid (Choi et al., 1991). The existence of a frequently used stable 'virus fold' was consistent with the concept that protein structures are composed of functionally useful conserved domains (Rossmann et al., 1974).

The structure of polyomavirus needs special mention (Rayment et al., 1982) as it showed major deviations from the classical Caspar–Klug $T = 7$ structure. However, the effective T number was found to be 6, which does not obey the Caspar–Klug rules. The rules had been accepted with such complete faith that it took a tremendous effort (Rayment, 1983) to convince others that the structure is correct and does not fully agree with the previous dogma. All this was later confirmed by higher-resolution X-ray results (Liddington et al., 1991) and by cryo-electron microscopy (Baker et al., 1989). Gradually, also, it was recognized that virus capsids have many important functions. For instance, many plant viruses swell in the absence of divalent cations like Ca^{2+}, which may be required for uncoating when the virus enters the cytoplasm. Harrison showed that swelling of TBSV was associated with some remarkable rotations of the whole capsid protein (Robinson & Harrison, 1982).

The first spherical virus structures to be solved were all plant viruses as it was easy to obtain these in large quantity given a simple greenhouse. Studying animal viruses was extremely difficult because it required, in general, cell culture with which only relatively minute quantities of virus could be grown. However, the first animal virus structure, namely that of a human common cold virus (human rhinovirus 14, HRV14), was solved by us in 1985 (Rossmann et al., 1985) at 3.5 Å resolution, enough for placing most of the atoms in space, followed a few months later by that of poliovirus (Hogle et al., 1985). These structures were both a major advance in phase determination and in providing biological insight. In the HRV14 structure, phases were extended from 6 to 3.5 Å resolution by using the power of non-crystallographic averaging, a procedure that was thought by

many to be impossible. The structure showed where neutralizing antibodies bind and suggested what might be required for neutralization; it hinted (correctly) at the site of binding of the cellular receptor and showed how the stages of assembly direct the assembly process; and, finally, the structure of the virus was seen to be related to the $T = 3$ plant viruses and, hence, of a common evolutionary origin. A year later, it was also shown where a series of antiviral compounds bind to HRVs and gave insight into their mechanism (Smith *et al.*, 1986).

Another feature of the HRV14 structure determination was the use of synchrotron radiation to obtain faster and better results. This allowed for rapid data collection where one data set required perhaps only one or two week's work (today less than one day) instead of a year of house X-ray usage. With these developments, the stage was set for an explosion of new and interesting structures, including the observation of ordered nucleic acid (Chen *et al.*, 1989; Tsao *et al.*, 1991; Larson *et al.*, 1993).

Undoubtedly, the greatest achievement is the structure determination of bluetongue virus (BTV) cores by David Stuart and his colleagues (Grimes *et al.*, 1997). BTV belongs to the group of double-stranded RNA segmented reoviruses. They contain two capsids, one inside the other separated by a large spacer protein. Most remarkable is that the internal capsid also contains polymerases which transcribe the viral RNA. This virus is about 800 Å in diameter, and cell dimensions of the actively transcribing virus are as big as 1600 Å. The structure determination has proceeded in steps with separate characterization of some of the structural proteins, the viral core (lacking the outside shell) and the whole virus. The cores show the RNA organization and are likely to show much of the transcriptional machinery.

Electron microscopy has been making an increasingly stronger impact on virus structure. Extending the ideas of DeRosier & Klug (1968), Crowther (1971) had shown how the projection data of various images could be combined to form a three-dimensional image. However, with negative staining, as was the practice prior to about 1987, such three-dimensional images were of only about 30 Å resolution. These were amazingly good and often rather informative, but there was a big gap between the power of X-ray crystallography and electron microscopy. This situation changed with the introduction of cryo-electron microscopy. Soon, numerous ~20 Å resolution structures started to appear, including that of Sindbis virus (Fuller, 1987), polyomavirus and SV40 (Baker *et al.*, 1989) and rotavirus (Prasad *et al.*, 1990). The big advantage here is that this method does not require crystals and does not even require a particularly pure sample as individual images of the correct object can be selected from the electron micrograph.

Further advances now include field-emission guns for the generation of electrons, permitting greater beam coherence and, hence, higher-resolution data provided suitable precautions are taken to consider where the phase-contrast function has its nodes and flips the phase (to view images with sufficient contrast requires that the beam is focused slightly above or below the specimen). These improvements have led to structures at 10 and 7 Å resolution of hepatitis B cores (Böttcher *et al.*, 1997; Conway *et al.*, 1997) in which helical structure can be discerned and followed.

9. The future

X-ray crystallography may soon reach its limits. It may be that assemblies much larger than BTV cannot have the capacity for order across the 1000 Å or more diameter, although, if there is such a limit, it has not yet been reached. However, the structures of larger biological assemblies are made to interact with their environment. Complexes of viruses with receptors, assembly intermediates containing scaffolding proteins, complexes of viruses with antibodies and viral fusion with cellular membranes are of major interest. Such complexes may be of rather transient duration and might not, therefore, be crystallizable. The obvious new tool for such structure is high-resolution cryo-electron microscopy. Similarly, in reconstructing electron microscopy images of viruses, it may be possible to orient each image to align the possibly structured nucleic acid and, hence, obtain the complete structure of a packaged genome. Thus, a combination of X-ray crystallography and even higher resolution and better cryo-electron microscopy is likely to be the way of study in the next decade or two. Recent examples of such combinations are BTV (Hewat *et al.*, 1994; Grimes *et al.*, 1997), Ross River virus (Cheng *et al.*, 1995) and the procapsid of φX174 (Dokland *et al.*, 1997). Use of electron-microscopy reconstructions will also allow the study of asymmetric objects or only partially symmetric objects, such as tailed bacteriophages (Tao *et al.*, 1998) as well as ribosomes (Yonath *et al.*, 1987; Yonath & Franceschi, 1993; Beckmann *et al.*, 1997). Thus, there may continue to be a rapid development of technique to study ever more interesting biological phenomena.

written for a volume of the *International Tables* on macromolecular structure being prepared for the International Union of Crystallography.

References

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). *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). *Nature (London)*, **227**, 1098–1103.

Arndt, U. W. & Wonacott, A. J. (1977). *The Rotation Method in Crystallography*. Amsterdam: North-Holland.

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). *Internationale Tabellen zur Bestimmung von Kristallstrukturen*, Vol. 1, edited by C. Hermann, pp. 84–87. Berlin: Begrüder Borntraeger.

Baker, T. S., Drak, J. & Bina, M. (1989). *Biophys. J.* **55**, 243–253.

Beckmann, R., Bubeck, D., Grassucci, R., Penczek, P., Verschoor, A., Blobel, G. & Frank, J. (1997). *Science*, **278**, 2123–2126.

Beevers, C. A. & Lipson, H. (1934). *Philos. Mag.* **17**, 855–859.

Berman, H. & Sussman, J. L. (1998). *Nature Struct. Biol.* In the press.

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

Bernal, J. D. & Crowfoot, D. (1934). *Nature (London)*, **133**, 794–795.

Bernal, J. D. & Fankuchen, I. (1941). *J. Gen. Physiol.* **25**, 111–165.

Bernal, J. D., Fankuchen, I. & Perutz, M. (1938). *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). *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). *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). *Nature (London)*, **206**, 757–761.

Blake, C. C. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1967). *Proc. R. Soc. London Ser. B*, **167**, 365–377.

Bloomer, A. C., Champness, J. N., Bricogne, G., Staden, R. & Klug, A. (1978). *Nature (London)*, **276**, 362–368.

Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.

Blow, D. M. & Rossmann, M. G. (1961). *Acta Cryst.* **14**, 1195–1202.

Bluhm, M. M., Bodo, G., Dintzis, H. M. & Kendrew, J. C. (1958). *Proc. R. Soc. London Ser. A*, **246**, 369–389.

Böttcher, B., Wynne, S. A. & Crowther, R. A. (1997). *Nature (London)*, **386**, 88–91.

Boyes-Watson, J., Davidson, E. & Perutz, M. F. (1947). *Proc. R. Soc. London Ser. A*, **191**, 83–132.

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

Bragg, W. H. (1915). *Philos. Trans. R. Soc. London Ser. A*, **215**, 253–274.

Bragg, W. H. & Bragg, W. L. (1913). *Proc. R. Soc. London Ser. A*, **88**, 428–438.

Bragg, W. L. (1913). *Proc. R. Soc. London Ser. A*, **89**, 248–277.

Bragg, W. L. (1929a). *Proc. R. Soc. London Ser. A*, **123**, 537–559.

Bragg, W. L. (1929b). *Z. Kristallogr.* **70**, 475–492.

Bragg, W. L. (1958). *Acta Cryst.* **11**, 70–75.

Bricogne, G. (1976). *Acta Cryst.* A**32**, 832–847.

Buehner, M., Ford, G. C., Moras, D., Olsen, K. W. & Rossmann, M. G. (1974). *J. Mol. Biol.* **82**, 563–585.

Burnett, R. M., Grütter, M. G. & White, J. L. (1985). *J. Mol. Biol.* **185**, 105–123.

Butler, P. J. G. & Klug, A. (1978). *Sci. Am.* **239**(5), 62–69.

Caspar, D. L. D. (1956). *Nature (London)*, **177**, 475–476.

Caspar, D. L. D. (1963). *Adv. Protein Chem.* **18**, 37–121.

Caspar, D. L. D. & Klug, A. (1962). *Cold Spring Harbor Symp. Quant. Biol.* **27**, 1–24.

Chen, Z., Stauffacher, C., Li, Y., Schmidt, T., Bomu, W., Kamer, G., Shanks, M., Lomonossoff, G. & Johnson, J. E. (1989). *Science*, **245**, 154–159.

Cheng, R. H., Kuhn, R. J., Olson, N. H., Rossmann, M. G., Choi, H. K., Smith, T. J. & Baker, T. S. (1995). *Cell*, **80**, 1–20.

Choi, H. K., Tong, L., Minor, W., Dumas, P., Boege, U., Rossmann, M. G. & Wengler, G. (1991). *Nature (London)*, **354**, 37–43.

Cochran, W., Crick, F. H. C. & Vand, V. (1952). *Acta Cryst.* **5**, 581–586.

Conway, J. F., Cheng, N., Zlotnick, A., Wingfield, P. T., Stahl, S. J. & Steven, A. C. (1997). *Nature (London)*, **386**, 91–94.

Crowfoot, D. (1948). *Annu. Rev. Biochem.* **17**, 115–146.

Crowfoot, D., Bunn, C. W., Rogers-Low, B. W. & Turner-Jones, A. (1949). *The Chemistry of Penicillin*, edited by H. T. Clarke, J. R. Johnson & R. Robinson, pp. 310–367. Princeton University Press.

Crowfoot, D. & Schmidt, G. M. J. (1945). *Nature (London)*, **155**, 504–505.

Crowther, R. A. (1969). *Acta Cryst.* B**25**, 2571–2580.

Crowther, R. A. (1971). *Philos. Trans. R. Soc. London Ser. B*, **261**, 221–230.

Crowther, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. Rossmann, pp. 173–178. New York: Gordon and Breach.

Cullis, A. F., Muirhead, H., Perutz, M. F., Rossmann, M. G. & North, A. C. T. (1962). *Proc. R. Soc. London Ser. A*, **265**, 161–187.

Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. (1985). *Nature (London)*, **318**, 618–624.

DeRosier, D. J. & Klug, A. (1968). *Nature (London)*, **217**, 130–134.

Dickerson, R. E., Kendrew, J. C. & Strandberg, B. E. (1961). *Acta Cryst.* **14**, 1188–1195.

Dickerson, R. E., Takano, T., Eisenberg, D., Kallai, O. B., Samson, L., Cooper, A. & Margoliash, E. (1971). *J. Biol. Chem.* **246**, 1511–1535.

Dodson, E., Harding, M. M., Hodgkin, D. C. & Rossmann, M. G. (1966). *J. Mol. Biol.* **16**, 227–241.

Dokland, T., McKenna, R., Ilag, L. L., Bowman, B. R., Incardona, N. L., Fane, B. A. & Rossmann, M. G. (1997). *Nature (London)*, **389**, 308–313.

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

Franks, A. (1955). *Proc. Phys. Soc. London Ser. B*, **68**, 1054–1064.

Fuller, S. D. (1987). *Cell*, **48**, 923–934.

Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). *Proc. R. Soc. London Ser. A*, **225**, 287–307.

Grimes, J. M., Jakana, J., Ghosh, M., Basak, A. K., Roy, P., Chiu, W., Stuart, D. I. & Prasad, B. V. V. (1997). *Structure*, **5**, 885–893.

Hahn, T. (1983). Editor. *International Tables for Crystallography*, Vol. A, *Space-Group Symmetry*. Dordrecht: Reidel Publishing Co. (Present distributor Kluwer Academic Publishers, Dordrecht.)

Harker, D. (1956). *Acta Cryst.* **9**, 1–9.

Harker, D. & Kasper, J. S. (1947). *J. Chem. Phys.* **15**, 882–884.

Harrison, S. C. (1968). *J. Appl. Cryst.* **1**, 84–90.

Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. & Bricogne, G. (1978). *Nature (London)*, **276**, 368–373.

Hendrickson, W. A. (1991). *Science*, **254**, 51–58.

Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.

Henry, N. F. M. & Lonsdale, K. (1952). Editors. *International Tables for X-ray Crystallography*, Vol. I, *Symmetry Groups*. Birmingham, England: Kynoch Press.

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

Hewat, E. A., Booth, T. F. & Roy, P. (1994). *J. Struct. Biol.* **112**, 183–191.

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). *Proc. R. Soc. London Ser. A*, **242**, 228–263.

Hogle, J. M., Chow, M. & Filman, D. J. (1985). *Science*, **229**, 1358–1365.

Holm, L. & Sander, C. (1997). *Structure*, **5**, 165–171.

Holmes, K. C., Stubbs, G. J., Mandelkow, E. & Gallwitz, U. (1975). *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 & Co.]

Johnson, J. E. (1978). *Acta Cryst.* B**34**, 576–577.

Jones, T. A. (1978). *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). *Nature (London)*, **213**, 862–865.

Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. & Phillips, D. C. (1958). *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). *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). *Science*, **179**, 285–288.

Larson, S. B., Koszelak, S., Day, J., Greenwood, A., Dodds, J. A. & McPherson, A. (1993). *J. Mol. Biol.* **231**, 375–391.

Liddington, R. C., Yan, Y., Moulai, J., Sahli, R., Benjamin, T. L. & Harrison, S. C. (1991). *Nature (London)*, **354**, 278–284.

Liljas, L., Unge, T., Jones, T. A., Fridborg, K., Lövgren, S., Skoglund, U. & Strandberg, B. (1982). *J. Mol. Biol.* **159**, 93–108.

Linderström-Lang, K. U. & Schellman, J. A. (1959). *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.

Lonsdale, K. (1928). *Nature (London)*, **122**, 810.

Lonsdale, K. (1936). *Structure Factor Tables*. London: G. Bell.

McLachlan, D. Jr & Champaygne, E. F. (1946). *J. Appl. Phys.* **17**, 1006–1014.

Main, P. & Rossmann, M. G. (1966). *Acta Cryst.* **21**, 67–72.

Matthews, B. W. (1966). *Acta Cryst.* **20**, 230–239.

Matthews, B. W., Sigler, P. B., Henderson, R. & Blow, D. M. (1967). *Nature (London)*, **214**, 652–656.

Murzin, A. G., Brenner, S. E., Hubbard, T. & Chothia, C. (1995). *J. Mol. Biol.* **247**, 536–540.

North, A. C. T. (1965). *Acta Cryst.* **18**, 212–216.

Olby, R. (1974). *The Path to the Double Helix*. Seattle: University of Washington Press.

Patterson, A. L. (1934). *Phys. Rev.* **46**, 372–376.

Patterson, A. L. (1935). *Z. Kristallogr.* **90**, 517–542.

Pauling, L. & Corey, R. B. (1951). *Proc. Natl Acad. Sci. USA*, **37**, 729–740.

Pauling, L., Corey, R. B. & Branson, H. R. (1951). *Proc. Natl Acad. Sci. USA*, **37**, 205–211.

Pepinsky, R. (1947). *J. Appl. Phys.* **18**, 601–604.

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

Perutz, M. F. (1949). *Proc. R. Soc. London Ser. A*, **195**, 474–499.

Perutz, M. F. (1951a). *Nature (London)*, **167**, 1053–1054.

Perutz, M. F. (1951b). *Nature (London)*, **168**, 653–654.

Perutz, M. F. (1954). *Proc. R. Soc. London Ser. A*, **225**, 264–286.

Perutz, M. F. (1956). *Acta Cryst.* **9**, 867–873.

Perutz, M. G. (1962). *Proteins and Nucleic Acids. Structure and Function*. Amsterdam: Elsevier.

Perutz, M. (1985). *Methods Enzymol.* **114**, 3–18.

Perutz, M. F. (1997). *Selected Topics in the History of Biochemistry: Personal Reflections*, Vol. 5, edited by G. Semenza & R. Jaenicke, pp. 57–67. Amsterdam: Elsevier.

Ponder, J. W. & Richards, F. M. (1987). *J. Mol. Biol.* **193**, 775–791.

Prasad, B. V. V., Burns, J. W., Marietta, E., Estes, M. K. & Chiu, W. (1990). *Nature (London)*, **343**, 476–479.

Price, W. C. (1946). *Am. J. Bot.* **33**, 45–54.

Ramachandran, G. N. & Sasisekharan, V. (1968). *Adv. Protein Chem.* **23**, 283–438.

Rayment, I. (1983). *Acta Cryst.* A**39**, 102–116.

Rayment, I., Baker, T. S., Caspar, D. L. D. & Murakami, W. T. (1982). *Nature (London)*, **295**, 110–115.

Reeke, G. N., Hartsuck, J. A., Ludwig, M. L., Quiocho, F. A., Steitz, T. A. & Lipscomb, W. N. (1967). *Proc. Natl Acad. Sci. USA*, **58**, 2220–2226.

Richards, F. M. (1968). *J. Mol. Biol.* **37**, 225–230.

Robertson, J. M. (1935). *J. Chem. Soc.* pp. 615–621.

Robertson, J. M. (1936). *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). *Nature (London)*, **250**, 546–551.

Robinson, I. K. & Harrison, S. C. (1982). *Nature (London)*, **297**, 563–568.

Rossmann, M. G. (1960). *Acta Cryst.* **13**, 221–226.

Rossmann, M. G. (1961). *Acta Cryst.* **14**, 383–388.

Rossmann, M. G. (1972). *The Molecular Replacement Method.* New York: Gordon and 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). *Nature (London)*, **317**, 145–153.

Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.

Rossmann, M. G. & Blow, D. M. (1963). *Acta Cryst.* **16**, 39–45.

Rossmann, M. G., Moras, D. & Olsen, K. W. (1974). *Nature (London)*, **250**, 194–199.

Sanger, F. & Thompson, E. O. P. (1953a). *Biochem. J.* **53**, 353–366.

Sanger, F. & Thompson, E. O. P. (1953b). *Biochem. J.* **53**, 366–374.

Sanger, F. & Tuppy, H. (1951). *Biochem. J.* **49**, 481–490.

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

Smith, T. J., Kremer, M. J., Luo, M., Vriend, G., Arnold, E., Kamer, G., Rossmann, M. G., McKinlay, M. A., Diana, G. D. & Otto, M. J. (1986). *Science*, **233**, 1286–1293.

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

Tao, Y., Olson, N. H., Xu, W., Anderson, D. L., Rossmann, M. G. & Baker, T. S. (1998). *Cell.* In the press.

Tsao, J., Chapman, M. S., Agbandje, M., Keller, W., Smith, K., Wu, H., Luo, M., Smith, T. J., Rossmann, M. G., Compans, R. W. & Parrish, C. R. (1991). *Science*, **251**, 1456–1464.

Valegård, K., Liljas, L., Fridborg, K. & Unge, T. (1990). *Nature (London)*, **345**, 36–41.

Wang, B. C. (1985). *Methods Enzymol.* **115**, 90–112.

Watson, J. D. (1954). *Biochim. Biophys. Acta*, **13**, 10–19.

Watson, J. D. (1968). *The Double Helix.* New York: Atheneum.

Watson, J. D. & Crick, F. H. C. (1953a). *Nature (London)*, **171**, 737–738.

Watson, J. D. & Crick, F. H. C. (1953b). *Nature (London)*, **171**, 964–967.

Wyckoff, H. W., Hardman, K. D., Allewell, N. M., Inagami, T., Johnson, L. N. & Richards, F. M. (1967). *J. Biol. Chem.* **242**, 3984–3988.

Yonath, A. & Franceschi, F. (1993). *Curr. Opin. Struct. Biol.* **3**, 45–49.

Yonath, A., Leonard, K. R. & Wittmann, H. G. (1987). *Science*, **236**, 813-816.