

Evolution of diffraction methods for solving crystal structures

Wayne A. Hendrickson

Departments of Biochemistry and Molecular Biophysics and of Physiology and Cellular Biophysics, Columbia University, New York, NY 10032, USA. Correspondence e-mail:

wayne@xtl.cumc.columbia.edu

Received 10 December 2012

Accepted 13 December 2012

The discovery of X-ray diffraction in 1912 by Laue and co-workers had important implications for the physics of diffraction, for the nature of X-radiation and for the structure of matter. Lawrence Bragg made important contributions to early developments in each of these areas, but the most pregnant of his innovations was in structure determination from X-ray diffraction data. He continued to make highly significant contributions to structure determination right on to the first crystal structures of proteins. Crystallography has made substantial contributions to chemistry and biology, and notably so for biological macromolecules.

1. Introduction

The discovery of X-ray diffraction, announced one hundred years ago in celebrated papers by Friedrich, Knipping & Laue (1912) and Laue (1912), elicited nearly instantaneous appreciation and widespread reactions. Implications for the physics of the diffraction phenomenon, for the nature of X-rays and for understanding of the structure of matter were evident from the outset. W. Lawrence Bragg contributed significantly from the start in pursuing all of these implications, but it was through structure determinations that he had the greatest influence.

Laue's intuition that X-ray waves, having short wavelengths commensurate with spacings in crystals, might generate optical diffraction proved true; and his original treatment of the three-dimensional diffraction problem is that used to this day. With reference to the reciprocal lattice ($\mathbf{S} = h\mathbf{a}^* + k\mathbf{b}^* + l\mathbf{c}^*$) and the sphere of reflection introduced by Ewald (1913), the Laue conditions for diffraction ($\mathbf{S} \cdot \mathbf{a} = h$; $\mathbf{S} \cdot \mathbf{b} = k$; $\mathbf{S} \cdot \mathbf{c} = l$) define the diffraction from an extensive crystal. Laue's explanation (Laue, 1912) in terms of five different wavelengths, λ , proved to be wrong, however; and it remained but a short time for the first paper from Bragg (1913*a*, read 11 November 1912) to realize the correct explanation. Bragg appealingly described each diffracted ray in simple one-dimensional terms as a 'reflection' at the glancing angle θ from planes of atoms separated by the spacing d ; thus $n\lambda = 2d \sin \theta$, albeit cast initially as the cosine of the supplementary angle to the plane normal. The verity of Bragg's law was sealed with his demonstrations of diffraction from cleavage planes from mica (W. L. Bragg, 1912), and the formula correctly described Laue's results as the crystal-selected diffraction of wavelengths meeting both Bragg's law and its equivalents in Laue conditions.

Several other papers on the physics of X-ray diffraction followed quickly. Notably, Friedel (1913) established that diffraction imposes a center of symmetry (Friedel's law), while nevertheless distinguishing 11 symmetry groups, now known as Laue groups; Debye (1913) described how temperature affects the motion of atoms, showing that diffraction pertains to dynamics as well as to static structure; and Darwin (1914*a,b*) elaborated a comprehensive theory for the absolute intensities of diffraction both in the kinematic approximation for ideally imperfect (mosaic) crystals and in the dynamic case of repeated scattering in perfect crystals.

Although Laue left open the question of the wave *versus* particle (photon) character of X-rays, and others including W. H. Bragg (1912) had offered particle-based explanations for Laue's diffraction observations, W. L. Bragg's argument (Bragg, 1913*a*) explicitly ascribed the phenomenon to electromagnetic waves, *i.e.* light as we know it today.

Arguably the most profound of the implications of the Laue discovery was that for atomic structure. Here again, it was Bragg who led the way with his structure determinations of alkali halides (Bragg, 1913*b*), notably NaCl. Indeed, his prior analysis of zincblende (ZnS), based on Laue's results, had already come close to describing its structure (Bragg, 1913*a*). Many other structures followed, initially those of other minerals and salts, and later inorganic complexes and organic compounds of increasing complexity. Ultimately, the impact of structure has been the greatest for chemistry and biology with numbers of marvelous triumphs. In this paper, I summarize the evolution of methods for structure determination from the perspective of macromolecular crystallography, but paying homage to the small-molecule tradition from which methods for macromolecules have evolved. Schwarzenbach (2012) has given an account of the developments from the perspective of small-molecule crystallography.

2. Early approaches to structure determination

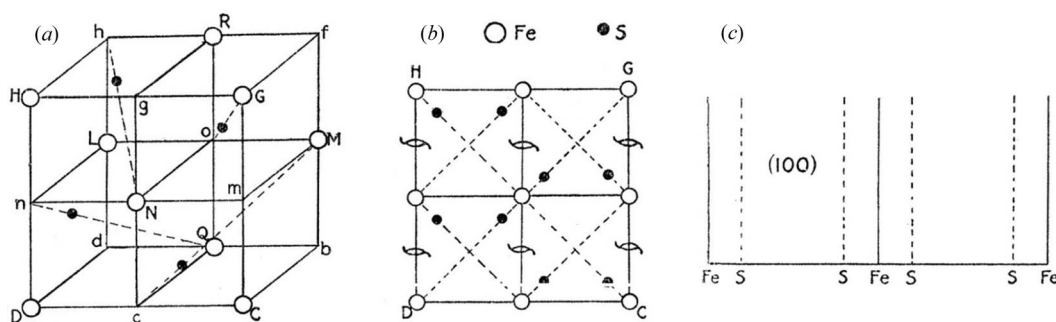
Bragg's approach to crystal structure determination for the series of alkali halides (Bragg, 1913a) was based on considerations of symmetry in the diffracting space lattices and the scattering strengths of the differing atoms (which he provisionally equated to atomic weight rather than electron count), and on unit-cell distances, which he cleverly calibrated by determining the X-ray wavelength (here 1.10 Å from Pt $L\beta$) from knowledge of molecular weight and crystal density. He considered rocksalt (NaCl) as the most interesting, needing two interpenetrating face-centered cubic arrays to explain its pattern of intensities, more complicated than that from KCl, where the two ions scatter equally, and those from KBr and KI, where the halide ion dominates. His father and he similarly determined the structure of diamond (Bragg & Bragg, 1913); and, using his father's X-ray spectrometer, he soon afterwards solved the structures of several other salts and minerals (Bragg, 1914), including zincblende (ZnS), fluorspar (CaF₂), iron pyrite (FeS₂) and calcite (CaCO₃). Whereas other early structures had all atomic positions fixed by symmetry, the sulfur atoms of pyrite and the oxygen atoms of calcite each depend on one free parameter, which Bragg determined by the comparison of observed intensity patterns with ones calculated from alternative models. A portion of Bragg's perceptive analysis for iron pyrite is illustrated in Fig. 1. This approach came to be known as trial-and-error.

Other early structure determinations also used the trial-and-error method, often employing algebraic or graphical methods to solve for variable parameters; however, such approaches became less tenable as the number of free parameters increased. On the other hand, once a structure that matched the measured diffraction data could be found, the

phases (actually signs, since all early structures were centrosymmetric) could be calculated and then Fourier images could be made. The Fourier representation of X-ray diffraction results was anticipated by W. H. Bragg (1915), but without practical elaboration. Others had made such calculations along one-dimensional lines of projection, but it was W. L. Bragg who first implemented Fourier analysis in two-dimensional projection, producing stunning images (Bragg, 1929; Fig. 2a). For this development, Bragg used the structure of diopside [CaMg(SiO₃)₂], dependent on 14 atomic parameters determined in the earlier structure determination (Warren & Bragg, 1928). For these calculations, he adopted a Fourier formulation by Ewald (1921).

Fourier analysis opened up many new avenues for structure determination, and prominent among these was the Patterson function (Patterson, 1934, 1935), which provided the means for comprehensive visualization of structural implications from diffraction patterns without prior assumptions or phase evaluations. Importantly for practical use, Harker (1936) realized that specific parameters could be read out from selected sections of three-dimensional Patterson functions. Many other developments ensued, and Patterson methods played a major role in small-molecule crystallography for many years, often accompanied by iterative Fourier analyses to complete partial structural interpretations.

The phase problem brought into light by Fourier analysis also opened up new, and ultimately more powerful, approaches to structure determination. An early glimpse of this prospect came in studies of the alums [KAl(SO₄)₂ hydrate and its analogs]. Cork (1927) showed that isomorphous series could be formed with NH₄, K, Rb or Tl substituting for Mg or with Cr substituting for Al; and Beevers & Lipson (1935) showed that Se could substitute isomorphously for S. Actual imple-



(d)

$l_{111}:l_{222}:l_{333}:l_{444}:l_{555}$ Ratios:
vs.
 $d(\text{Fe-S}):d(111)$ Ratios

$d(110)/d(111) = \sqrt{2}/\sqrt{3}$

Therefore, we have the following comparison:—

Ratio.	Intensities.
1 : 3	100 56 38 0 8
1 : 4	78 100 35 2 21
1 : 5	19 100 24 7 17
Observed intensities.....	75 100 50 0 22

The 1 : 4 ratio seems, therefore, to be very nearly the true one.

Figure 1

Bragg's solution of the structure of iron pyrite. (a) Front half of the face-centered cubic structure of iron pyrite. (b) Projection of the iron pyrite structure into the (100) face. (c) Alignment of Fe and S positions as projected into the (100) face. (d) Analysis of the sulfur position by spectral analysis of $I(hhh)$ intensities. Comparison of calculated and observed ratios of intensities (right) for the (hhh) reflections identified on the left. Similar comparisons were made for the $I(h00)$ and $I(hh0)$ spectral orders. Parts (a), (b) and (c) are adapted from Figs. 4 A and B of Bragg (1914), and the right-hand part of (d) was excised from the text (p. 484).

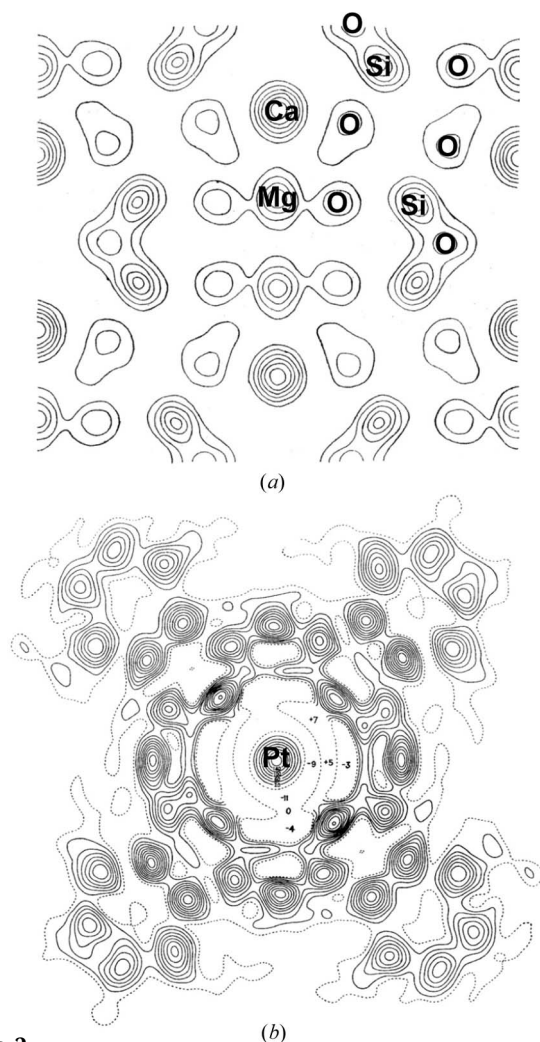


Figure 2 Electron-density distributions computed by Fourier synthesis. (a) Projection into the (001) direction of diopside $[\text{CaMg}(\text{SiO}_3)_2]$, adapted and annotated from Fig. 3A of Bragg (1929). (b) Projection into the (010) direction of platinum phthalocyanine, reproduced from Fig. 1 of Robertson & Woodward (1940).

mentation of isomorphous replacement in phase determination awaited work by Robertson (1936) on phthalocyanines, however. By comparing diffraction patterns from metal-free and the nickel-complexed phthalocyanine, and deducing that the metal site is at the cell origin, Robertson was able to determine signs for apo phthalocyanine in projection and Robertson & Woodward (1937) did so as well for nickel phthalocyanine. In this latter study, it was already noted that nickel dominated in the phasing such that most signs were positive. This suggested that phases from heavy metals might approximate those from the whole structure, and the effectiveness of this heavy-atom method was demonstrated influentially in the structure of platinum phthalocyanine (Robertson & Woodward, 1940; Fig. 2b). With respect to isomorphous replacement, a critical development was the extension to non-centrosymmetric structures, which came in a series of studies by Bijvoet *et al.* (1948, 1949) that culminated in the crystal structure of strychnine (Bokhoven *et al.*, 1951) based on the substitution of sulfate by selenate. With these

developments of heavy-atom methods, the path to determination of arbitrarily complex structures was set.

As structure determinations moved to focus on the phase problem and Fourier analysis, attention to mathematical approaches also increased. It was evident that the number of diffraction observations from a typical crystal often greatly exceeds the number of variable parameters, whereby the system of intensity equations should specify the variables, albeit in a manner frustrated by non-linearity. Similarly, this recognized overdetermination motivated the expectation of relations among the structure factors, *i.e.* phase relationships. Notable among the early mathematical developments were the inequalities of Harker & Kasper (1948) based on algebraic conditions, the determinants of Karle & Hauptman (1950) based on the constraint of positivity, and Sayre's equation (Sayre, 1952) based on the condition of atomicity. With recognition of implications from probability (Hauptman & Karle, 1953; Karle & Hauptman, 1956), the triplet relationship that $\varphi(\mathbf{h}) \simeq \varphi(\mathbf{k}) + \varphi(\mathbf{h} - \mathbf{k})$ was seen to be central to effective phase determination. Here $\mathbf{h}(hkl)$ is a reciprocal-lattice vector and the relationship holds provided that the associated structure-factor amplitudes $|F(\mathbf{h})|$, $|F(\mathbf{k})|$ and $|F(\mathbf{h} - \mathbf{k})|$ are sufficiently strong. Computer implementations of resulting direct methods came to dominate in small-molecule crystallography.

In parallel with these later developments, an appreciation for the potential of anomalous scattering in structural analysis also emerged. Bijvoet (1949) suggested that anomalous-scattering effects could be used in phase determination, and with co-workers he used such effects in a ground-breaking determination of the absolute configuration of tartaric acid; using a crystal of Na Rb tartrate and radiation above the Rb absorption edge, he tested for deviations $[|F(\mathbf{h})| \neq |F(-\mathbf{h})|]$ from Friedel's law (Bijvoet *et al.*, 1951). Various formulations and experimental tests of anomalous scattering for phase determination followed (Okaya *et al.*, 1955; Okaya & Pepinsky, 1956; Ramachandran & Raman, 1956; Ramaseshan & Venkatesan, 1957); however, substantial impact on structure analysis did not emerge until much later.

3. Diffraction-constrained model building

Just as Lawrence Bragg combined understanding of atomic principles with constraints from diffraction to solve the first structures of salts and minerals, so too were some of the earliest and most influential of structures for biological macromolecules devised by model building constrained by diffraction experiments. Having learned detailed atomic parameters from crystal structures of amino acids and dipeptides and knowing of Astbury's classifications of α - and β -proteins based on fiber diffraction (Astbury & Street, 1932), Pauling and co-workers used first principles of hydrogen bonding to discover the structures of the α -helix (Pauling *et al.*, 1951) and of β -pleated sheets (Pauling & Corey, 1951). Realizing that Pauling's model of the α -helix made better sense than the helical structures he had built with Bragg and Kendrew (Bragg *et al.*, 1950), Perutz verified the Pauling

model by inclining fibers of poly γ -benzyl-L-glutamate and α -keratin (horse hair) until the axial reflection expected from the 1.5 Å/residue repeat would fall on Ewald's sphere (Perutz, 1951). Pauling had ignored the mismatch of the strong 5.15 Å axial reflection from keratin-family fibers (Astbury & Street, 1932) with the 5.4 Å/turn on an α -helix. Crick, on the other hand, chose to explain the discrepancy. Accepting that the α -helix must be true, he inclined helices by 18° and twisted them about the fiber axis with knobs-into-holes packing (Crick, 1952); thereby, he also invented what we now know to be a highly prevalent folding pattern for proteins. Then, in a most marvelous stroke of genius and insight, Watson & Crick (1953) discovered the double-helical structure of DNA, taking advantage of Franklin's observation of a 3.4 Å axial repeat in DNA fibers (Franklin & Gosling, 1953).

4. Biological macromolecules and the small-molecule tradition

Although significant elements of proteins and nucleic acids may have regularly repeating structures, as was modeled by Pauling and co-workers for the α -helices and β -pleated sheets of proteins and by Watson and Crick for B-form DNA, biological macromolecules are in general far too asymmetric and complicated for structure analysis by model-building approaches. The richness of diffraction patterns from protein crystals demonstrated that such X-ray data should suffice to specify the three-dimensional structure (Bernal & Crowfoot, 1934); however, structure determination clearly would require the measurement of intensities for complete diffraction patterns, not just for projection data, and also the evaluation of general phases, not just signs as for centrosymmetric structures. The heavy-atom method was having success with quite large structures, such as in Hodgkin's structural analysis of vitamin B₁₂ (Hodgkin *et al.*, 1956), and direct methods came to solve large all-light-atom structures, *e.g.* α -amanitin (Karle *et al.*, 1973); however, even small proteins are an order of magnitude larger than these. It was not initially clear that isomorphous replacement would be powerful enough either, but ultimately this proved successful for hemoglobin, first for a centrosymmetric projection (Green *et al.*, 1954) and then for a non-centrosymmetric projection (Blow, 1958) where three heavy-atom derivatives were used. The full three-dimensional structure of myoglobin, first at 6 Å resolution (Kendrew *et al.*, 1958) and later at 2 Å resolution (Kendrew *et al.*, 1960) made it clear that protein structures could be solved by the method of multiple isomorphous replacement (MIR), and ultimately the structure of hemoglobin also succumbed into a structure in three dimensions at 2.8 Å resolution (Perutz *et al.*, 1968).

Macromolecular crystallography has borrowed much more than isomorphous replacement from the small-molecule tradition. While Patterson functions of macromolecules are too complex to be solved, $(\Delta F)^2$ -difference Pattersons gave excellent approximations for solving the substructures of heavy atoms needed for the isomorphous-replacement method. Later, direct methods were adapted to this role, again using difference coefficients. While the heavy-atom method

itself is usually inapplicable to macromolecules, it is the essence of the molecular replacement method, which has become a mainstay. Fourier series are commonly used iteratively, just as for earlier small molecules, to improve on partial models fitted to initial maps. While anomalous scattering played a limited role for small molecules, it came to provide powerfully complementary information to augment MIR phasing into MIR with anomalous scattering (MIRAS) and to resolve ambiguities from single isomorphous replacement (SIR) as the SIRAS method (Blow & Rossmann, 1961; Herriott *et al.*, 1970). These and other adaptations from the small-molecule tradition required many improvements and innovations. Importantly, the combination of phase information from multiple sources was optimized by the development of phase probability distributions (Blow & Crick, 1959; Hendrickson & Lattman, 1970). All said, a process emerged with considerable maturity as recorded by Blundell & Johnson (1976) in the first text on protein crystallography. By this time, holdings of the Protein Data Bank (PDB; Berman *et al.*, 2000) had reached 76 deposits (15 October 1976) and some other structures had not been deposited. Only 13 PDB deposits from 1976 remain today; others were made obsolete by refined replacements (see below).

5. Changing practice in macromolecular crystallography

Maturity did not mean stagnation for macromolecular crystallography. To the contrary, several subsequent innovations went well beyond the small-molecule tradition, and thereby largely transformed the practice of how macromolecular structures are solved from X-ray diffraction. Transformative innovations came in structure refinement, molecular replacement, density modification and anomalous diffraction; moreover, technological advances have synergized to accelerate the changes.

By 1976, most structures were simply presented as a model built by hand to fit the density; with the development of stereochemically restrained refinement (Hendrickson & Konnert, 1980; Konnert & Hendrickson, 1980) and combined X-ray and energy refinement (Brünger, 1991), structures were improved by bringing them into optimal agreement with their diffraction patterns.

With few exceptions, the structures accumulated by 1976 had been determined *de novo* by MIR or SIR, although some were isomorphous follow-on complexes with relevant ligands. The rotation function (Rossmann & Blow, 1962) had given the possibility to detect structural similarities within or between crystals, and by now structures were accumulating such that an unknown macromolecule of interest might be related to an already known structure and be expected to be similar. The method of molecular replacement, a phase borrowing closely related to that of the heavy-atom method, was innovated to determine unknown structures from knowns or known portions (Lattman *et al.*, 1971; Rossmann, 1972), and this method now dominates as the number of known structures to serve as templates has grown very large.

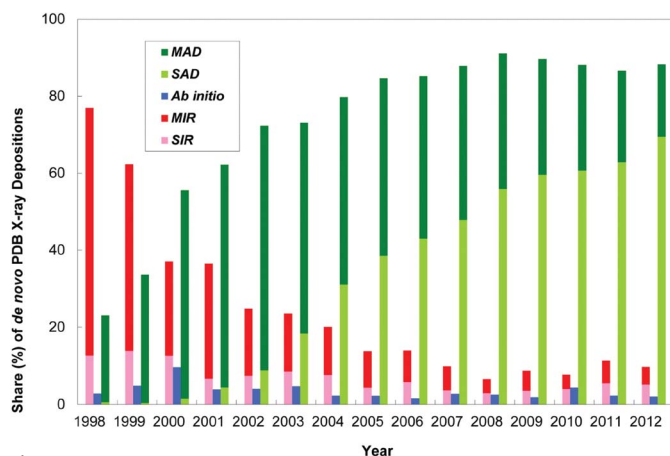


Figure 3

Changing practice in *de novo* structure determination. Depositor declarations for the method of structure determination were parsed from PDB entries beginning in 1998, the first full year for such recordings. Numbers were normalized into percentages of all *de novo* determinations in each year, and plotted as histograms as identified by color coding. Multiple declarations (e.g. SAD and MIR or MAD and molecular replacement) were counted for each method, but there were few such instances. The MIR numbers include MIRAS determinations and SIR includes SIRAS; *ab initio* numbers also include 'direct methods' and related declarations. Results are based on the PDB holdings of 2 October 2012.

That non-crystallographic symmetry had implications for phase evaluation (Rossmann & Blow, 1962) and could be used in phase improvement (Rossmann & Blow, 1963; Crowther, 1967; Bricogne, 1974) was well understood by 1976, but it had not yet played a major role in structure determinations. That soon changed dramatically with molecular averaging used in atomic level structures of viruses (Harrison *et al.*, 1978; Abad-Zapatero *et al.*, 1980) and viral proteins (Champness *et al.*, 1976; Wilson *et al.*, 1981; Varghese *et al.*, 1983). Many macromolecular structures possess symmetry, but all have solvent expanses, and more generalized density-modification methods (Wang, 1985; Cowtan & Main, 1993), which use solvent flattening, histogram matching and related modifications for phase improvement, found broad application. Moreover, automated chain-tracing procedures (Perrakis *et al.*, 1999), which hybridize density modification with structure refinement, do truly automate many macromolecular structure determinations.

Anomalous scattering had figured in many MIR or SIR structures by 1976, as MIRAS or SIRAS, but it was not yet then a prime player. That changed when the structure of crambin was solved from the anomalous scattering of its native sulfur atoms (Hendrickson & Teeter, 1981), which we can see in retrospect as the first application of the method of single-wavelength anomalous diffraction (SAD). Ultimately, with the availability of tunable X-rays from synchrotron sources, practical implementation of the method of multiwavelength anomalous diffraction (MAD) was realized (Kahn *et al.*, 1985; Hendrickson, 1985; Fourme & Hendrickson, 1990; Hendrickson, 1991) and first applied to previously unknown structures shortly afterwards (Guss *et al.*, 1988; Hendrickson *et al.*, 1989). This approach blossomed, fueled importantly

by the incorporation of selenomethionine into proteins (Hendrickson *et al.*, 1990). By 2000, MAD had overtaken MIR and SIR combined, and by 2006, with advances in density-modification procedures (Wang, 1985; Cowtan & Main, 1993) for resolving the phase ambiguity inherent in its single-wavelength counterpart, SAD had overtaken MAD for dominance in *de novo* structure determination (Fig. 3). Currently, nearly 70% of *de novo* macromolecular structures are determined by SAD alone; and, whereas in 1996 over 80% were determined by MIR and SIR combined, since 2007 nearly 90% are done by the MAD and SAD alternatives. Recently, through the use of multiple crystals to enhance signal-to-noise for the weak anomalous diffraction from light atoms ($Z \leq 20$) in macromolecules, SAD analysis directly from native biological macromolecules appears to be robustly feasible (Liu *et al.*, 2012).

A thorough description of technological advances would go beyond the scope of this review, but it is important to give notice to some of the prime developments. Obviously prominent among these are computer technology, including interactive computer graphics (Jones, 1985), and synchrotron radiation, which was brought to the attention of crystallographers earlier (Rosenbaum *et al.*, 1971) but did not have significant impact until the 1980s. Remarkably, the brightness of X-ray sources has increased even faster than for Moore's law of computing power; and the latest development of X-ray free-electron lasers promises great opportunities in nanocrystallography (Boutet *et al.*, 2012). The remarkable advances in molecular biology that flowed from the structure of DNA have also been crucial for progress in macromolecular crystallography. The benefits for crystallographers include the ready production of recombinant proteins as identified in the complete genomic sequences of many organisms and the facility for easily producing recombinant selenomethyl proteins for structure analysis and mutant variants to test structure-inspired hypotheses.

6. Triumphs of macromolecular crystallography

The fruits of macromolecular crystallography are resplendent by many measures – by the number of structures that have been determined, by the size and complexity of structures that can be analyzed, and by the recognized significance of many of these resulting structures.

The active holdings of the PDB (Berman *et al.*, 2000) as of 4 December 2012 were 86 661 released deposits, of which 88.0% are X-ray crystal structures (<http://www.rcsb.org/pdb/>). These deposits include 8483 from 2012 alone, a number that already nearly matches the total accumulation through to 1998. Besides these structures of biological macromolecules, the stores of information unleashed by Bragg's development of crystal structure determination include hundreds of thousands of additional structures in the Cambridge Structural Database of organic and metallo-organic compounds (<http://www.ccdc.cam.ac.uk/products/csd/>), in the Inorganic Crystal Structure Database ([Acta Cryst. \(2013\). A69, 51–59](http://www.fiz-</p>
</div>
<div data-bbox=)

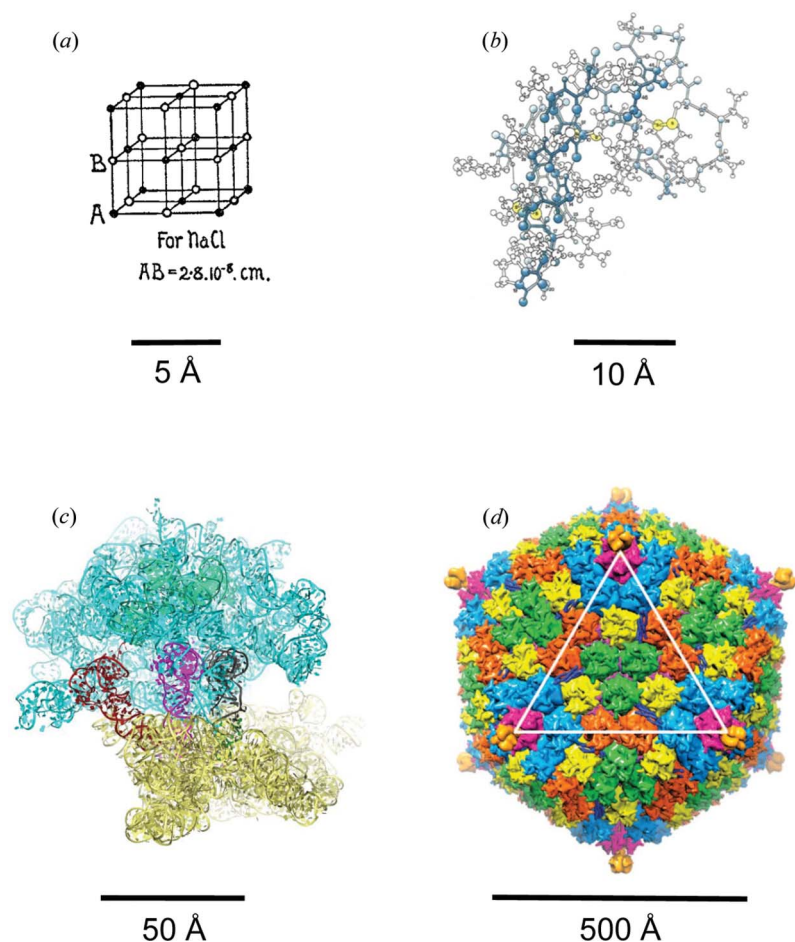


Figure 4

Progressive increases in the complexity of crystal structures. Scales for linear dimensions are as labeled on associated bars. (a) Rocksalt (NaCl). This is a crystal in space group $Fm\bar{3}m$ with lattice constant $a = 5.640 \text{ \AA}$; thus, its unit-cell volume (V_{uc}) is 180 \AA^3 and the volume of the asymmetric unit (V_{au}) is $V_{uc}/192 = 0.93 \text{ \AA}^3$. The image is adapted from Fig. 10 of Bragg (1913b) where A and B may represent Na and Cl, respectively, or *vice versa*. (b) Crambin. This protein molecule is from a crystal in space group $P2_1$ with $a = 40.96$, $b = 18.65$, $c = 22.52 \text{ \AA}$ and $\beta = 90.77^\circ$; thus its $V_{au} = V_{uc}/2 = 8601 \text{ \AA}^3$. The image here was adapted from a drawing made by Irving Geis based on the coordinates of Hendrickson & Teeter (1981). All atoms are shown, including hydrogen atoms. The six sulfur atoms, from which the structure was determined by what is now known as SAD phasing, are in yellow. (c) Complex of a 70S ribosome with an mRNA and three tRNAs. This particle is from a crystal in space group $P2_12_12_1$ with $a = 213.3$, $b = 453.0$ and $c = 631.4 \text{ \AA}$; thus its $V_{au} = V_{uc}/4 = 15\,252\,240 \text{ \AA}^3$ for the two ribosome particles per asymmetric unit. The image is reproduced from Fig. S1 of Selmer *et al.* (2006). Reprinted with permission from AAAS. The image shows only the RNA portions from one ribosome (50S bluish, 30S yellowish and tRNAs brown, magenta and dark gray). Phosphate backbones are drawn as ribbons, and ribose and base rings are drawn as colored planes. The 54 ordered protein molecules of this 70S ribosomal structure are omitted from the figure. (d) Human adenovirus capsid. This virus structure is from a crystal in space group $P1$ with lattice constants $a = 854.0$, $b = 855.2$, $c = 865.2 \text{ \AA}$, $\alpha = 119.6$, $\beta = 91.7$ and $\gamma = 118.1^\circ$; thus its $V_{au} = V_{uc} = 450\,275\,173 \text{ \AA}^3$ for one viral particle. The image is reproduced from Fig. 1A of Reddy *et al.* (2010). Reprinted with permission from AAAS. It shows a surface rendering viewed down a threefold axis of the icosahedral particle. One of 20 identical facets is circumscribed by the white triangle, which has three of the 12 viral penton proteins at its vertices and 12 of the 240 viral hexon proteins uniquely within its boundaries. The icosahedral asymmetric unit thus comprises one-fifth of a vertex penton (pink) and four hexons (one cyan, one orange, one yellow and one green) plus accessory cement proteins (deep blue, between orange and yellow hexons; magenta, between hexons at four places) and a trimeric fiber (yellow–orange) at each penton base. Other accessory proteins are at the capsid interior and out of view.

karlsruhe.de/icsd_content.html), and in the CRYSTMET (<http://www.tothcanada.com/>) database of metals and alloys.

Each structure in the PDB is large, typically comprising several thousand atoms and, with time, ever increasingly intricate molecules and assemblages are being solved. With Fourier-based methods of analysis, the size of structure is limited essentially only by the degree of order that can be achieved and typically not by technical challenges. The increasing complexity of crystal structures is illustrated in Fig. 4, where Bragg's rocksalt structure (Bragg, 1913b) is compared with structures of crambin (Hendrickson & Teeter, 1981), among the smallest of proteins, and two recent large structures, a 70S ribosome–tRNA–mRNA complex (Selmer *et al.*, 2006) and a human adenovirus particle (Reddy *et al.*, 2010). These structures cover a range of linear dimensions of nearly two orders of magnitude, as indicated by scales as shown, and a range of 5×10^8 in the volumes of crystallographic asymmetric units, as listed in the figure legend.

With rare exception, each of the thousands of macromolecular structures has provided insights into the biochemical action of these molecules, and oftentimes the clues to biological function have been revolutionary for the associated science. The first atomic level structures of proteins provided insights into the principles of protein structure and chemistry, but from the outset they also had impact on biochemistry and medicine. It is not possible to do justice to the many contributions of crystallography to biology and chemistry in this brief account, and I can only apologize for missing many superb examples, but it would be another injustice not to note some highlights.

The first structures of myoglobin (Kendrew *et al.*, 1960) and hemoglobin (Perutz *et al.*, 1968), and following analyses, gave insights into oxygen transport. Early structures of enzymes, typified by lysozyme (Blake *et al.*, 1965) and chymotrypsin (Matthews *et al.*, 1967), provided clear pictures of the catalytic process, and many enzyme structures followed; nowadays, enzymology is essentially a crystallographic science. Important insights into immune recognition emerged from initial studies on immunoglobulin fragments (Poljak *et al.*, 1973) and a histocompatibility antigen (Bjorkman *et al.*, 1987), and great numbers have followed. Applications to viruses and viral proteins, cited above in connection with molecular averaging, have had many ramifications including membrane fusion and drugs for infectious diseases.

The first membrane protein structure, that of a photosynthetic reaction center (Deisenhofer *et*

al., 1985), was followed by many important electron-transfer complexes involved in photosynthesis and respiration. Atomic level crystal structures of transcription factors in complexes with target DNA segments, first in bacterial systems (Aggarwal *et al.*, 1988; Otwinowski *et al.*, 1988) and later for the basal transcription initiation complex of eukaryotes (J. L. Kim *et al.*, 1993; Y. Kim *et al.*, 1993), were but the start of an incredibly productive enterprise. The HIV pandemic led to several significant structural projects including its protease (Navia *et al.*, 1989) and reverse transcriptase (Kohlstaedt *et al.*, 1992), which facilitated drug discoveries that can keep the disease in check, and its envelope glycoprotein (Kwong *et al.*, 1998), which has inspired a program for vaccine development. The structures of actin (Kabsch *et al.*, 1990) and myosin (Rayment *et al.*, 1993) gave initial atomic substance to the machinery of muscle contraction. The structures of Hsp70 heat-shock molecular chaperone components (Flaherty *et al.*, 1990; Zhu *et al.*, 1996) and of a bacterial chaperonin (Braig *et al.*, 1994) opened the field of chaperone-assisted protein folding to structural definition. Structures of protein kinases, first for a serine/threonine kinase (Knighton *et al.*, 1991) and then for the catalytic domain of a receptor tyrosine kinase (Hubbard *et al.*, 1994) paved the way for drug-discovery projects across the pharmaceutical industry. The structure of the F1 ATP synthase (Abrahams *et al.*, 1994) provided a framework for understanding the catalytic machinery for proton-gradient fueled synthesis of ATP.

The structure of a potassium channel (Doyle *et al.*, 1998) began a continuing elucidation of ion-conduction channels of profound significance for neurobiology. The structure of the nucleosome core particle (Luger *et al.*, 1997) provided definition for the fundamental unit of chromosome organization and a basis for understanding the impact of epigenetic modification. The structure of the ribosome at the atomic level of resolution, first for its 30S (Wimberly *et al.*, 2000; Schluenzen *et al.*, 2000) and 50S components (Ban *et al.*, 2000), and later for intact 70S particles (Schuwirth *et al.*, 2005) including associated mRNA and tRNA factors (Selmer *et al.*, 2006), provided detailed snapshots of the basic translational machinery. Similarly, the structure of yeast RNA polymerase II (Cramer *et al.*, 2001) and following developments gave insights into the molecular mechanism of eukaryotic transcription of genetic information in DNA into messenger RNA (mRNA) for translation into proteins. Most recently, the structures of G-protein coupled receptors (GPCRs) have been determined, first for bovine rhodopsin in its inactive state (Palczewski *et al.*, 2000) and most recently for an activated human β -adrenergic receptor in complex with an agonist and its cognate heterotrimeric G-protein (Rasmussen *et al.*, 2011).

This small selection of noteworthy contributions from macromolecular crystallography includes a number that have also been recognized in major honors to associated investigators. Adding to the Nobel Prize awarded to the Braggs for initiating structure determination and several other such awards for crystallographic investigations, a total of seven Nobel Prizes in Chemistry have recognized macromolecular crystal structures that are cited in this section as being a

significant basis for those awards. These Prizes include four of the past ten in chemistry.

I thank Arno Pähler for a program to parse the PDB to identify the method used in structure determination and Helen Berman for the historical record of PDB holdings in 1976.

References

- Abad-Zapatero, C., Abdel-Meguid, S. S., Johnson, J. E., Leslie, A. G., Rayment, I., Rossmann, M. G., Suck, D. & Tsukihara, T. (1980). *Nature (London)*, **286**, 33–39.
- Abrahams, J. P., Leslie, A. G., Lutter, R. & Walker, J. E. (1994). *Nature (London)*, **370**, 621–628.
- Aggarwal, A. K., Rodgers, D. W., Drott, M., Ptashne, M. & Harrison, S. C. (1988). *Science*, **242**, 899–907.
- Astbury, W. T. & Street, A. (1932). *Philos. Trans. R. Soc. A*, **230**, 75–101.
- Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. (2000). *Science*, **289**, 905–920.
- Beevers, C. A. & Lipson, H. (1935). *Proc. R. Soc. A*, **148**, 664–680.
- Berman, H., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. & Bourne, P. E. (2000). *Nucl. Acids Res.* **28**, 235–242.
- Bernal, J. D. & Crowfoot, D. (1934). *Nature (London)*, **133**, 794–795.
- Bijvoet, J. M. (1949). *Proc. Acad. Sci. Amst. B*, **52**, 313–314.
- Bijvoet, J. M., Bokhoven, C. & Schoone, J. C. (1948). *K. Ned. Akad. Wet. Proc.* **51**, 990.
- Bijvoet, J. M., Bokhoven, C. & Schoone, J. C. (1949). *K. Ned. Akad. Wet. Proc.* **52**, 120–121.
- Bijvoet, J. M., Peerdeman, A. F. & van Bommel, A. J. (1951). *Nature (London)*, **168**, 271–272.
- Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L. & Wiley, D. C. (1987). *Nature (London)*, **329**, 506–512.
- Blake, C. C., Koenig, D. F., Mair, G. A., North, A. C., Phillips, D. C. & Sarma, V. R. (1965). *Nature (London)*, **206**, 757–761.
- Blow, D. M. (1958). *Proc. R. Soc. A*, **247**, 302–336.
- 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.
- Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*. London: Academic Press.
- Bokhoven, C., Schoone, J. C. & Bijvoet, J. M. (1951). *Acta Cryst.* **4**, 275–280.
- Boutet, S., *et al.* (2012). *Science*, **337**, 362–364.
- Bragg, L., Kendrew, J. C. & Perutz, M. F. (1950). *Proc. R. Soc. A*, **203**, 321–357.
- Bragg, W. H. (1912). *Nature (London)*, **90**, 219.
- Bragg, W. H. (1915). *Philos. Trans. R. Soc. A*, **215**, 243–274.
- Bragg, W. H. & Bragg, W. L. (1913). *Proc. R. Soc. A*, **89**, 277–291.
- Bragg, W. L. (1912). *Nature (London)*, **90**, 410.
- Bragg, W. L. (1913a). *Proc. Camb. Philos. Soc.* **17**, 43–57.
- Bragg, W. L. (1913b). *Proc. R. Soc. A*, **89**, 248–277.
- Bragg, W. L. (1914). *Proc. R. Soc. A*, **89**, 468–489.
- Bragg, W. L. (1929). *Proc. R. Soc. A*, **123**, 537–559.
- Braig, K., Otwinowski, Z., Hegde, R., Boisvert, D. C., Joachimiak, A., Horwich, A. L. & Sigler, P. B. (1994). *Nature (London)*, **371**, 578–586.
- Bricogne, G. (1974). *Acta Cryst.* **A30**, 395–405.
- Brünger, A. T. (1991). *Annu. Rev. Phys. Chem.* **42**, 197–223.
- Champness, J. N., Bloomer, A. C., Bricogne, G., Butler, P. G. & Klug, A. (1976). *Nature (London)*, **259**, 20–24.
- Cork, J. M. (1927). *Philos. Mag.* **4**, 688–698.
- Cowtan, K. D. & Main, P. (1993). *Acta Cryst.* **D49**, 148–157.

- Cramer, P., Bushnell, D. A. & Kornberg, R. D. (2001). *Science*, **292**, 1863–1876.
- Crick, F. H. C. (1952). *Nature (London)*, **170**, 882–883.
- Crowther, R. A. (1967). *Acta Cryst.* **22**, 758–764.
- Darwin, C. (1914a). *Philos. Mag.* **27**, 315–333.
- Darwin, C. (1914b). *Philos. Mag.* **27**, 675–690.
- Debye, P. (1913). *Ann. Phys.* **43**, 49–95.
- Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. (1985). *Nature (London)*, **318**, 618–624.
- Doyle, D. A., Morais Cabral, J., Pfuetzner, R. A., Kuo, A., Gulbis, J. M., Cohen, S. L., Chait, B. T. & MacKinnon, R. (1998). *Science*, **280**, 69–77.
- Ewald, P. P. (1913). *Phys. Z.* **14**, 465–472.
- Ewald, P. P. (1921). *Z. Kristallogr.* **56**, 129–156.
- Flaherty, K. M., DeLuca-Flaherty, C. & McKay, D. B. (1990). *Nature (London)*, **346**, 623–628.
- Fourme, R. & Hendrickson, W. A. (1990). *Synchrotron Radiation and Biophysics*, edited by S. S. Hasnain, pp. 156–175. Chichester: Ellis Horwood.
- Franklin, R. E. & Gosling, R. G. (1953). *Nature (London)*, **171**, 740–741.
- Friedel, G. (1913). *C. R. Acad. Sci.* **157**, 1533–1536.
- Friedrich, W., Knipping, P. & Laue, M. (1912). *Sitz. Bayer. Akad. Wiss.* pp. 303–322.
- Green, D. W., Ingram, V. M. & Perutz, M. F. (1954). *Proc. R. Soc. A*, **225**, 287–307.
- Guss, J. M., Merritt, E. A., Phizackerley, R. P., Hedman, B., Murata, M., Hodgson, K. O. & Freeman, H. C. (1988). *Science*, **241**, 806–811.
- Harker, D. (1936). *J. Chem. Phys.* **4**, 381.
- Harker, D. & Kasper, J. S. (1948). *Acta Cryst.* **1**, 70–75.
- Harrison, S. C., Olson, A. J., Schutt, C. E., Winkler, F. K. & Bricogne, G. (1978). *Nature (London)*, **276**, 368–373.
- Hauptman, H. & Karle, J. (1953). *Solution of the Phase Problem I. The Centrosymmetric Crystal*. Am. Crystallogr. Assoc. Monograph No. 3. Polycrystal Book Service.
- Hendrickson, W. A. (1985). *Trans. Am. Crystallogr. Assn.*, **21**, 11–21.
- Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
- Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
- Hendrickson, W. A. & Konnert, J. H. (1980). *Computing in Crystallography*, edited by R. Diamond, S. Rameshan & K. Venkatesan, pp. 13.01–13.23. Bangalore: Indian Academy of Sciences.
- Hendrickson, W. A. & Lattman, E. E. (1970). *Acta Cryst.* **B26**, 136–143.
- Hendrickson, W. A., Pähler, A., Smith, J. L., Satow, Y., Merritt, E. A. & Phizackerley, R. P. (1989). *Proc. Natl Acad. Sci. USA*, **86**, 2190–2194.
- Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
- Herriott, J. R., Sieker, L. C., Jensen, L. H. & Lovenberg, W. (1970). *J. Mol. Biol.* **50**, 391–406.
- Hodgkin, D. C., Kamper, J., Mackay, M., Pickworth, J., Trueblood, K. N. & White, J. G. (1956). *Nature (London)*, **178**, 64–66.
- Hubbard, S. R., Wei, L., Ellis, L. & Hendrickson, W. A. (1994). *Nature (London)*, **372**, 746–754.
- Jones, T. A. (1985). *Methods Enzymol.* **115**, 157–171.
- Kabsch, W., Mannherz, H. G., Suck, D., Pai, E. F. & Holmes, K. C. (1990). *Nature (London)*, **347**, 37–44.
- Kahn, R., Fourme, R., Bosshard, R., Chiadmi, M., Risler, J. L., Dideberg, O. & Wery, J. P. (1985). *FEBS Lett.* **179**, 133–137.
- Karle, I. L., Karle, J., Wieland, T., Burgermeister, W., Faulstich, H. & Witkop, B. (1973). *Proc. Natl Acad. Sci. USA*, **70**, 1836–1840.
- Karle, J. & Hauptman, H. (1950). *Acta Cryst.* **3**, 181–187.
- Karle, J. & Hauptman, H. (1956). *Acta Cryst.* **9**, 635–651.
- 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, J. L., Nikolov, D. B. & Burley, S. K. (1993). *Nature (London)*, **365**, 520–527.
- Kim, Y., Geiger, J. H., Hahn, S. & Sigler, P. B. (1993). *Nature (London)*, **365**, 512–520.
- Knighton, D. R., Zheng, J. H., Ten Eyck, L. F., Ashford, V. A., Xuong, N. H., Taylor, S. S. & Sowadski, J. M. (1991). *Science*, **253**, 407–414.
- Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. & Steitz, T. A. (1992). *Science*, **256**, 1783–1790.
- Konnert, J. H. & Hendrickson, W. A. (1980). *Acta Cryst.* **A36**, 344–350.
- Kwong, P. D., Wyatt, R., Robinson, J., Sweet, R. W., Sodroski, J. & Hendrickson, W. A. (1998). *Nature (London)*, **393**, 648–659.
- Lattman, E. E., Nockolds, C. E., Kretsinger, R. H. & Love, W. E. (1971). *J. Mol. Biol.* **60**, 271–277.
- Laue, M. (1912). *Sitz. Bayer. Akad. Wiss.* pp. 363–373.
- Liu, Q., Dahmane, T., Zhang, Z., Assur, Z., Brasch, J., Shapiro, L., Mancina, F. & Hendrickson, W. A. (2012). *Science*, **336**, 1033–1037.
- Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. (1997). *Nature (London)*, **389**, 251–260.
- Matthews, B. W., Sigler, P. B., Henderson, R. & Blow, D. M. (1967). *Nature (London)*, **214**, 652–656.
- Navia, M. A., Fitzgerald, P. M., McKeever, B. M., Leu, C. T., Heimbach, J. C., Herber, W. K., Sigal, I. S., Darke, P. L. & Springer, J. P. (1989). *Nature (London)*, **337**, 615–620.
- Okaya, Y. & Pepinsky, R. (1956). *Phys. Rev.* **103**, 1645–1647.
- Okaya, Y., Saito, Y. & Pepinsky, R. (1955). *Phys. Rev.* **98**, 1857–1858.
- Otwinowski, Z., Schevitz, R. W., Zhang, R. G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F. & Sigler, P. B. (1988). *Nature (London)*, **335**, 321–329.
- Palczewski, K., Kumasaka, T., Hori, T., Behnke, C. A., Motoshima, H., Fox, B. A., Le Trong, I., Teller, D. C., Okada, T., Stenkamp, R. E., Yamamoto, M. & Miyano, M. (2000). *Science*, **289**, 739–745.
- Patterson, A. (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**, 251–256.
- Pauling, L., Corey, R. B. & Branson, H. R. (1951). *Proc. Natl Acad. Sci. USA*, **37**, 205–211.
- Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nat. Struct. Biol.* **6**, 458–463.
- Perutz, M. F. (1951). *Nature (London)*, **167**, 1053–1054.
- Perutz, M. F., Muirhead, H., Cox, J. M. & Goaman, L. C. (1968). *Nature (London)*, **219**, 131–139.
- Poljak, R. J., Amzel, L. M., Avey, H. P., Chen, B. L., Phizackerley, R. P. & Saul, F. (1973). *Proc. Natl Acad. Sci. USA*, **70**, 3305–3310.
- Ramachandran, G. N. & Raman, S. (1956). *Curr. Sci.* **25**, 348–351.
- Ramaseshan, S. & Venkatesan, K. (1957). *Curr. Sci.* **26**, 352–353.
- Rasmussen, S. G., et al. (2011). *Nature (London)*, **477**, 549–555.
- Rayment, I., Rypniewski, W. R., Schmidt-Bäse, K., Smith, R., Tomchick, D. R., Benning, M. M., Winkelmann, D. A., Wesenberg, G. & Holden, H. M. (1993). *Science*, **261**, 50–58.
- Reddy, V. S., Natchiar, S. K., Stewart, P. L. & Nemerow, G. R. (2010). *Science*, **329**, 1071–1075.
- Robertson, J. M. (1936). *J. Chem. Soc.* p. 1195.
- Robertson, J. M. & Woodward, I. (1937). *J. Chem. Soc.* p. 219.
- Robertson, J. M. & Woodward, I. (1940). *J. Chem. Soc.* p. 36.
- Rosenbaum, G., Holmes, K. C. & Witz, J. (1971). *Nature (London)*, **230**, 434–437.
- Rossmann, M. G. (1972). Editor. *The Molecular Replacement Method*. New York: Gordon & Breach.
- Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.
- Rossmann, M. G. & Blow, D. M. (1963). *Acta Cryst.* **16**, 39–45.
- Sayre, D. (1952). *Acta Cryst.* **5**, 60–65.

- Schluenzen, F., Tocilj, A., Zarivach, R., Harms, J., Gluehmann, M., Janell, D., Bashan, A., Bartels, H., Agmon, I., Franceschi, F. & Yonath, A. (2000). *Cell*, **102**, 615–623.
- Schuwirth, B. S., Borovinskaya, M. A., Hau, C. W., Zhang, W., Vila-Sanjurjo, A., Holton, J. M. & Cate, J. H. (2005). *Science*, **310**, 827–834.
- Schwarzenbach, D. (2012). *Acta Cryst.* **A68**, 57–67.
- Selmer, M., Dunham, C. M., Murphy, F. V., Weixlbaumer, A., Petry, S., Kelley, A. C., Weir, J. R. & Ramakrishnan, V. (2006). *Science*, **313**, 1935–1942.
- Varghese, J. N., Laver, W. G. & Colman, P. M. (1983). *Nature (London)*, **303**, 35–40.
- Wang, B. C. (1985). *Methods Enzymol.* **115**, 90–112.
- Warren, B. E. & Bragg, W. L. (1928). *Z. Kristallogr.* **56**, 129–156.
- Watson, J. D. & Crick, F. H. (1953). *Nature (London)*, **171**, 737–738.
- Wilson, I. A., Skehel, J. J. & Wiley, D. C. (1981). *Nature (London)*, **289**, 366–373.
- Wimberly, B. T., Brodersen, D. E., Clemons, W. M., Morgan-Warren, R. J., Carter, A. P., Vornrhein, C., Hartsch, T. & Ramakrishnan, V. (2000). *Nature (London)*, **407**, 327–339.
- Zhu, X., Zhao, X., Burkholder, W. F., Gragerov, A., Ogata, C. M., Gottesman, M. E. & Hendrickson, W. A. (1996). *Science*, **272**, 1606–1614.