Science Progress (2017), 100(1), 25–44
Paper 1600200 doi:10.3184/003685017X14858694684395

100 years of X-ray crystallography

TERENCE J. KEMP and NATHANIEL W. ALCOCK
Department of Chemistry, University of Warwick, Coventry, CV4 7AL, UK

Terence Kemp is an Emeritus Professor of Chemistry at the University of Warwick. He was educated at Cardiff High School, Watford Grammar School and Jesus College, Oxford. At Oxford he gained the degrees of MA, DPhil and DSc. After postdoctoral work at Leeds University on radiation chemistry, he joined the then new University of Warwick in 1965 for his remaining academic career. He has published 250 papers covering radiation chemistry, free radical solution kinetics, laser flash photolysis, ESR of (i) reactions of solvated electrons in liquid ammonia and (ii) solution and frozen redox photosystems, particularly of uranyl ion, metal complexes and diazonium salts, the latter yielding the triplet state aryl cation. His later interests covered semiconductor photocatalysis and polymer systems including linear polysulfides and nitrated polyethers. He is currently Editor of *Progress in Reaction Kinetics and Mechanism* and an Editor of *Science Progress.*
E-mail: T.J.Kemp@sciencereviews.co.uk

Nathaniel Alcock is an Emeritus Reader in Chemistry at the University of Warwick. He was educated at Trinity College, Cambridge, gaining the degrees of BA and PhD. After postdoctoral work in Oslo and Cambridge he joined the University of Warwick in 1966, lecturing on many aspects of inorganic chemistry. He has published more than 400 papers, principally on structural chemistry and secondary bonding and an overview of the subject (*Bonding and Structure: Structural Principles in Inorganic and Organic Chemistry,* 1990). He is also keenly interested in vernacular architecture and related aspects of local and social history, on which he has published extensively.
E-mail: n.w.alcock@warwick.ac.uk

ABSTRACT

The developments in crystallography, since it was first covered in Science Progress in 1917, following the formulation of the Bragg equation, are described. The advances in instrumentation and data analysis, coupled with the application of computational methods to data analysis, have enabled the solution of molecular structures from the simplest binary systems to the most complex of biological structures. These developments are shown to have had major impacts in the development of chemical bonding theory and in offering an increasing understanding of enzyme–substrate interactions. The advent of

synchrotron radiation sources has opened a new chapter in this multi-disciplinary field of science.

Keywords: *bonding, crystal structure, data analysis, DNA, electron density, powder diffraction, ribosome, symmetry, synchrotron, unit cell*

1. Introduction

1.1 Historical

The year 2017 is a rather suitable point to chart the developments in X-ray crystallography bearing in mind that the Nobel Prize in Physics in 1915 was awarded to the father and son team of William Henry Bragg and William Lawrence Bragg (Figure 1)[1]. It is significant that, since then, crystallography has been (and remains) one of the most multidisciplinary sciences that links together frontier areas of research and has, directly or indirectly, produced the largest number of Nobel Laureates throughout the history of the awards, with 29 Prizes for 48 Laureates up to the present day.

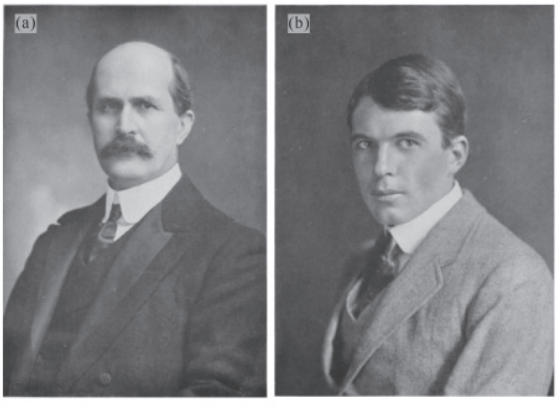


Figure 1 *(a) W.H. Bragg and (b) W.L Bragg from Science Progress[3].*

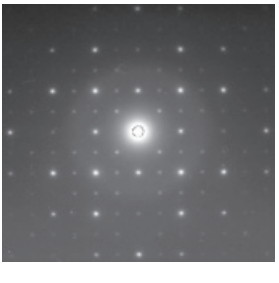


Figure 2 *An X-ray diffraction pattern from the single crystal of sodium chloride.*

The Braggs had built[2] on the discovery by Max von Laue of the phenomenon of the diffraction of X-rays in crystals, the latter having recognised that the wavelength of X-rays matched the spacing of atoms within solids[2]. The diffraction pattern from the NaCl crystal is shown in Figure 2. A report of the Braggs' award was made in Science Progress in 1916[3].

The initial discovery left some difficult problems in that not only the space lattices, but also the wavelengths and the intensity distribution wavelengths in the spectra of the X-rays, were unknown quantities. W.L. Bragg

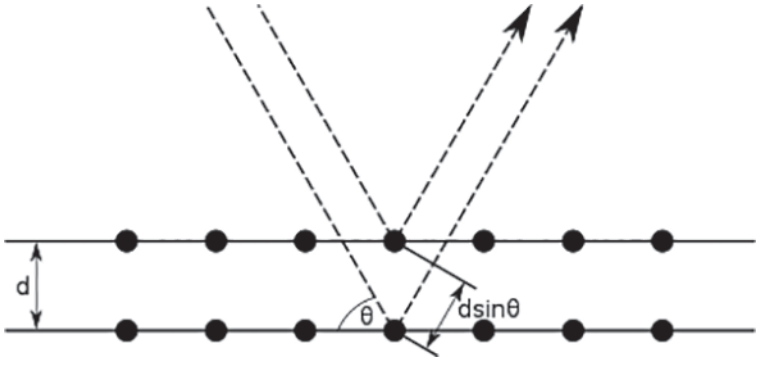


Figure 3 *Bragg diffraction: two beams, with identical wavelength and phase, approach a crystalline solid and are scattered off two different atoms within it. The lower beam traverses an extra length of 2d sin θ. Constructive interference occurs when this length is equal to an integral multiple of the wavelength of the radiation.*

found that the phenomenon could be treated mathematically as a reflection by the successive parallel planes that may be placed so as to pass through the lattice points (Figure 3). In this way, the ratio between the wavelengths and the distances of these planes from each other could be calculated from the angle of reflection. The Bragg formula (Bragg's law) can be written:

$$n\lambda = 2d \sin \theta \quad (1)$$

The Braggs focused their initial efforts on the simplest structures, *i.e.* the alkali metal halides, demonstrating the face-centred cubic lattice of NaCl, KBr and KI and the simple cubic lattice of CsCl. In the first, a metal atom is surrounded by six halide ions and each halide ions, and in the second structure each ion has eight neighbours. The Braggs' investigation of diamond revealed the tetrahedral nature of the environment of each carbon atom[1].

Since these original discoveries, X-ray crystallography has become one of the most important scientific techniques for determining the structure of molecules from the smallest to the largest. This article traces its development over the 100 years following the publication of Braggs' law.

2. Terminology

While it is not the aim here to enable readers to solve X-ray structures, it is necessary to introduce some of the terms and parameters used in X-ray diffraction.

2.1 Unit cell

A crystal lattice is built made up of identical 3-dimensional units which are repeated by translation in the X-, Y- and Z-directions. The unit cell (Figure 4) can be thought of as the fundamental structural pattern from which the crystal is constructed. Unit

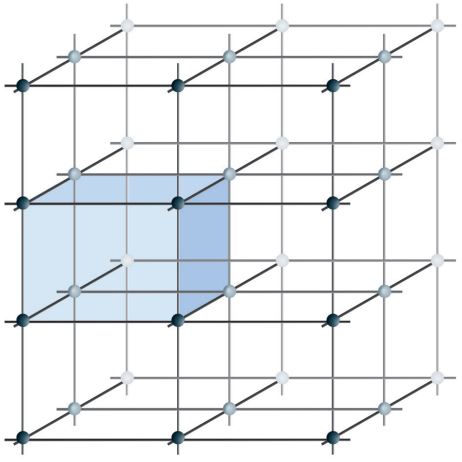


Figure 4 *A unit cell (shaded area) is a parallel-sided figure which, by translations in all directions, builds up the crystal lattice.*

cells are classified into seven crystal systems by noting the rotational symmetry elements they possess; thus a tetragonal unit cell has one four-fold axis and a cubic unit cell has four three-fold axes arranged tetrahedrally.

2.2 Scattering factors

The scattering of X-rays is a result of the oscillations an incoming electromagnetic wave generates in the electrons of atoms: in particular, heavy atoms give rise to more scattering than light atoms. The extent of scattering depends on the number of electrons in the atom concerned as measured by the scattering factor *f* of the element. More precisely, the scattering factor is related to the electron density distribution in the atom [ρ(*r*)], by

$$f = 4\pi \int_0^\infty \rho(\mathrm{r}) \frac{\sin kr}{kr} r^2 \, \mathrm{d}r \quad \text{where} \quad k = \frac{4\pi}{\lambda} \sin\theta \tag{2}$$

The value of *f* is greatest in the forward direction and smaller for directions away from the forward direction. Detailed analysis of the intensities of the reflections must take account of this dependence on direction. It can be shown that for the forward direction, when θ equals zero, *f* is proportional to the number of electrons on the atom, *i.e.* its atomic number *Z*, as shown in Figure 5.

2.3 The electron density

For structures containing several different atoms, it is necessary to go beyond Bragg's law. If a unit cell contains several atoms with scattering factors f_j and coordinates ($x_j a$, $y_j b$, $z_j c$) then it can be shown that the overall amplitude of a wave diffracted by the (*hkl*) planes is given by:

$$F_{hkl} = \sum_j f_j e^{i\phi_{hkl}(j)} \quad \text{where} \quad \phi_{hkl}(j) = 2\pi \left(hx_j + ky_j + lz_j \right) \tag{3}$$

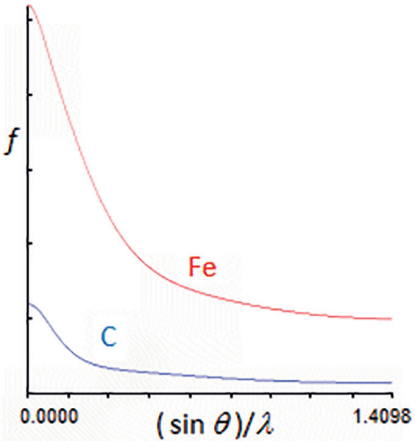


Figure 5 *Variation of scattering factors f for C and Fe atoms with atomic number and angle. At sin θ/λ = 0, f = Z. Ticks on the horizontal axis correspond to increments of 10° and ticks on the vertical axis correspond to increments of 5 electrons.*

The sum is over all the atoms in the unit cell and the quantity F_{hkl} is called the structure factor.

The intensity of the (*hkl*) reflection is proportional to $|F_{hkl}|^2$, so one can determine the amplitude of the structure factor by taking the square root of the corresponding intensities measured experimentally (*e.g.* by the blackening of a particular spot on a photographic film). If all structure factors are known, (*not only their amplitude but also their phase*), it is possible to calculate the electron density [ρ(x*yz*)] in the unit cell using Eqn (4)

$$\rho(xyz) = \frac{1}{V}\sum_{hkl} F_{hkl} e^{-2\pi i\,(hx+ky+lz)} \qquad (4)$$

where *V* is the volume of the unit cell. Mathematically, this equation is the inverse of Eqn (3). Summing the contribution from each of the structure factors (a Fourier synthesis) gives the electron density at each point in the unit cell. However, this procedure is at first sight of no practical use, since the measurement of the structure factors gives only their amplitudes not their phase and the latter is needed for the contributions from each structure factor to be combined correctly.

2.4 The phase problem

The structure factor can be viewed as a vector with an amplitude (length) and a direction in the X-Y plane (the phase angle). For mathematical convenience, the X-component of the vector is treated as 'real' and the Y-component as 'imaginary'. If the crystal contains a centre of inversion ('centre of symmetry'), the structure factor only has an X-component, so it is either positive (+X) or negative (–X).

The intensity enables one to determine $|F_{hkl}|$ but does not give information about its phase. This is known as the phase problem. Without knowing the phases of the structure factors, the electron density ρ(*xyz*) cannot be evaluated.

One method of overcoming the phase problem is the Patterson synthesis[4] which is particularly useful for molecules which contain one (or very few) heavier atoms combined with light atoms (C, N, O, H, *etc.*) such as metal–organic complexes. Here, instead of the structure factors F_{hkl}, the values of $|F_{hkl}|^2$, which can be obtained from the observed intensities, are used in an expression analogous to Eqn. (4)

$$P(xyz) = \frac{1}{V}\sum_{hkl} | F_{hkl} |^2 e^{-2\pi i\,(hx+ky+lz)} \qquad (5)$$

A Patterson synthesis gives us a map of the vector separations of the atoms in the unit cell, *i.e.* the distances and directions between atoms. Thus, if atom A is at coordinates (x_A, y_A, z_A) and atom B is at (x_B, y_B, y_B) then there will be a peak at $(x_A–x_B, y_A–y_B, z_A–z_B)$ in the Patterson map (there will also be a peak at the negative of these coordinates). Heavy atoms dominate F_{hkl} and so from the peaks in the Patterson map, the heavy atom positions can be determined. Using these positions to calculate the structure factors gives approximate phases, from which a Fourier synthesis should show both the heavy and light atom positions. Thus it is possible to gain knowledge of the geometry of an organic molecule by complexing it with a heavy metal ion as in the structure in Figure 6 for the complex between a uranyl ion, UO_2^{2+} and furoic acid.

Another approach is to introduce heavy atoms into crystals of a large molecule, *e.g.* of a protein, by soaking it in a solution of the heavy atom, or by co-crystallising with a salt of the heavy atom. This is termed isomorphous replacement. It assumes the structure of the protein is not subject to significant perturbation by this treatment. Heavy atoms that have been used in this way are Hg^+ (binding to thiol groups), Pb^{2+} (binding to cysteine residues) and $PtCl_4^-$ (binding to histidine).

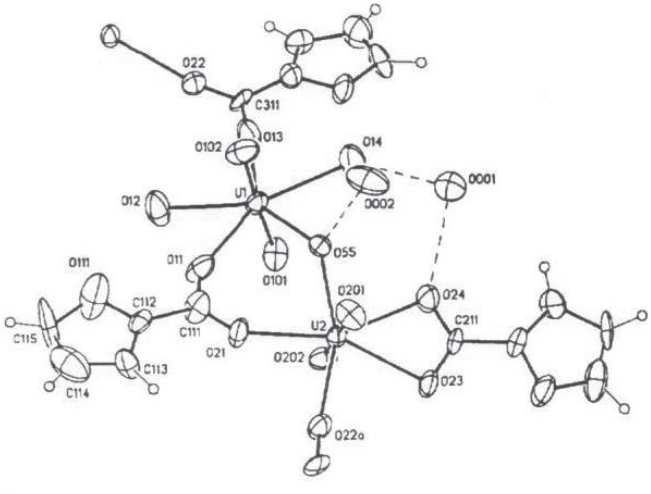


Figure 6 *View of the two distinct units of the uranyl–furoate complex, showing the numbering of the principal atoms*[5].

2.5 Development of direct methods

A major step forward was made in determining the structure of crystals containing only light atoms through the development of *direct methods* by Herbert A. Hauptman, a mathematician and Jerome Karle, a physicist, for which they were jointly awarded the Nobel Prize in Chemistry in 1985[6]. The direct methods depend on two facts. Firstly, the electron density, which diffracts the X-rays, can never be negative and secondly, the number of measurements is much greater than the number of parameters (the atomic positions) to be determined. This enables the application of statistical methods. In their work, done between 1950 and 1956, Hauptman and Karle laid the foundations for a rational exploitation of these possibilities, especially the use of inequalities. The advent of increasingly powerful computers facilitated the optimum utilisation of the Hauptman–Karle methodology.

2.6 Structure refinement

The aim here is to achieve the optimum fit between the observed intensities and those calculated from the model of the structure deduced from the diffraction pattern. This iterative process is termed structure refinement and now uses the 'least-squares' mathematical procedure (minimising the sum of the squares of the differences between the observed and calculated F values). The progress of refinement is measured by the R-factor, defined as

$$R = \frac{\sum |F_{obs} - F_{calc}|}{\sum |F_{obs}|} \qquad (6)$$

where F_{obs} and F_{calc} are the observed and calculated F values respectively.

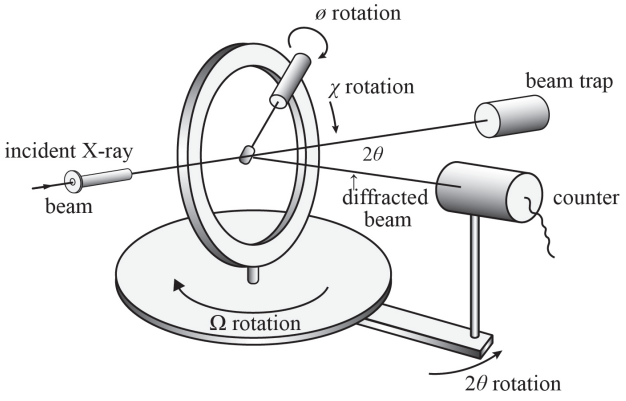
The R factor is calculated during each cycle of refinement to assess progress. The final R factor is a measure of the quality of the model. Small organic molecules commonly refine to $R < 0.05$; a desirable R factor for a protein molecule is *ca.* 0.2.

3. Data collection

3.1 Instrumentation

In a standard single crystal determination the crystal is mounted at the centre of an X-ray diffractometer on a goniometer head, which allows translations in three dimensions over a small range, so that the crystal can be placed precisely in the X-ray beam. The diffractometer is used to position the crystal at selected orientations. The crystal is illuminated with a finely focused beam of monochromatic X-rays and the scattered X-rays are detected on photographic film or, more recently, by an area detector with a charge-coupled device (CCD). The result is a diffraction pattern of the type shown in Figure 2, *i.e.* a series of regularly-spaced spots known as reflections. The two-dimensional images taken at different orientations are converted into the individual structure factor amplitudes (F_{obs}), which are then used in the structure factor determination. A typical modern instrument, a four-circle diffractometer, is shown in Figure 7.

(A)



(B)

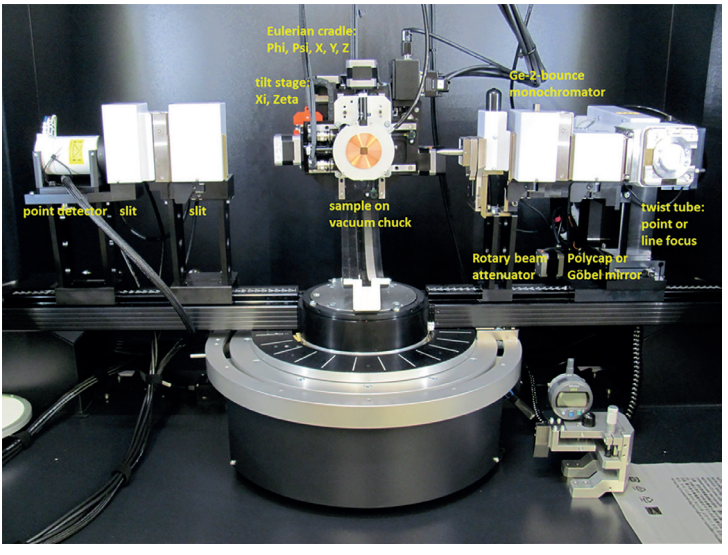


Figure 7 *Four-circle diffractometer: (A) schematic and (B) instrument with its principal components (Bruker Model D8 Discover).*

A single image of spots is insufficient to reconstruct the whole crystal as it will represent only a tiny fraction of the complete diffraction pattern. The crystal has to be rotated stepwise through 180° with an image being recorded at each incremental step. The task is then to determine (a) the *hkl* value for each spot (indexing), (b) the relative strengths of the spots in different images (merging and scaling) and (c) how the variations should be combined to yield the total electron density (phasing).

The first step in data processing begins with indexing the reflections. This involves determining the dimensions of the unit cell and which image peak corresponds to which position in reciprocal space. Indexing also leads to the symmetry of the crystal, *i.e.* its space group. The data from hundreds of images are then integrated, yielding a single list of indexed *F*-values.

3.2 Synchrotron radiation

The development of synchrotron sources has had a major effect on X-ray crystallography. The X-rays are generated in the synchrotron, when electrons are accelerated to near the speed of light within a storage ring. As the electron beam is bent by a magnet, the electrons emit X-rays. The energy produced by the X-ray beam means that the sample has to be cooled to 77 K to withstand thermal decomposition. A big advantage of this type of source is that it allows selection of the X-ray wavelength. It is particularly appropriate for examination of powder samples (see Section 3.4). It has also proved its utility in solving the structure of molecules such as the anti-inflammatory veterinary drug caprofen, which exists in polymorphic forms beyond the scope of conventional X-ray diffractometers (Figure 8). Another major benefit is the intensity of the X-ray beam, which shortens the time needed to collect sufficient numbers of reflections. The Diamond synchrotron source (www.diamond.ac.uk) is the major UK provider. It is located at the Harwell Science and Innovation Centre, Oxfordshire.

3.3 Neutron diffraction

A beam of neutrons is scattered by a crystal in the same way as a beam of X-rays and Bragg's law holds for both techniques. In neutron diffraction one measures the counts of scattered neutrons as a function of 2θ. The degree of scattering from an atom is controlled by its neutron scattering factor which, unlike the situation with X-rays, does not depend on the number of electrons around the nucleus (the atomic number) but on its nuclear structure. In particular, the scattering from H atoms is strong, in contrast to that for X-rays, which makes the technique especially important for structure determination of molecules featuring hydrogen bonding. More information about the potential of neutron diffraction can be found at the website of the UK facility ISIS (http://www.isis.stfc.ac.uk/instruments/neutron-diffraction2593.html).

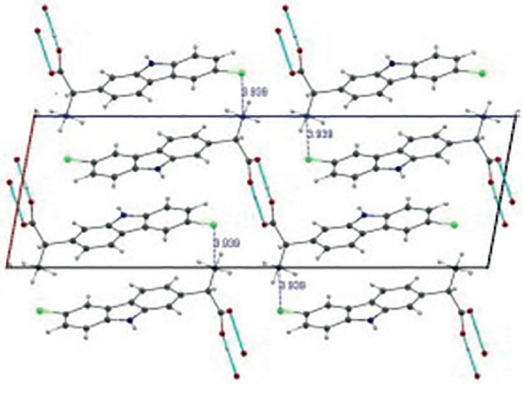


Figure 8 *Structure of caprofen as determined by powder X-ray method using synchrotron radiation*[7].

3.4 X-ray powder diffraction

A different approach to the study of the interaction of X-rays and crystals was taken by Debye and Scherrer[8] in 1916. They examined the diffraction of X-rays by a microcrystalline powder; here there are a vast number of orientations of the crystals and some of them will be in the correct one to meet the Bragg condition. In general, powder X-ray diffraction will yield less accurate structures compared to single-crystal determination and the latter will be the preferred method. However, it is not always the case that a single crystal of the target molecule can be obtained and so recourse has to be made to the powder technique. The advent of more powerful computational techniques has enabled the structure of molecules as complex as the anti-fungal drug griseofulvin (Figure 9) to be solved by X-ray powder diffraction[9].

Powder X-ray diffraction works particularly well when a synchrotron source provides the X-rays (Section 3.2).

3.5 Use of computers in crystallography

The application of electronic computers to X-ray crystallography from the mid-1960s onwards (50 years after the first use of the technique) transformed structure determination. In particular, it allowed the calculation of 3-dimensional electron density maps (by Fourier synthesis), refinement by least squares and (rather later) the development of direct methods of phase determination.

Previous to this, the immense number of manual calculations needed restricted the use of Fourier syntheses to two-dimensional maps and refinement was generally undertaken from 'difference' electron-density maps, showing how the observed density differed from that calculated for the approximately known atomic positions. In these maps, a negative region indicates that the atom should be removed, while a positive region suggests a position for a new atom.

The pioneers of the technique, Arnold Beevers and Henry Lipson, first showed the value of Fourier syntheses in 1934 in determining the structure of $CuSO_4.5H_2O$. However, even these 2-dimensional maps are very onerous to calculate until these

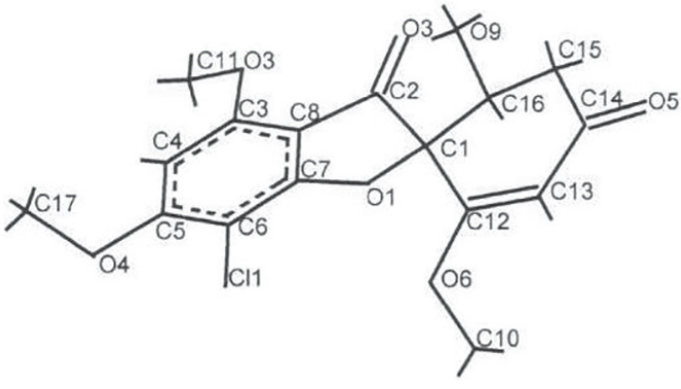


Figure 9 *Optimised structure of griseofulvin from X-ray powder diffraction[9].*

authors developed the very influential Beevers–Lipson strips which made the calculation of the maps more feasible[10].

4. The move to larger molecules

4.1 Proteins and vitamins

Improvements in experimentation and data analysis and, in the later 20th century, the development of direct methods and particularly the advent of computer systems capable of both collecting and carrying out data analysis, led to the solving of the structures of molecules such as vitamins and proteins. The first structures of very large molecules were solved successfully in the late 1950s, for example those of globular proteins such as sperm whale myoglobin by Kendrew and that of haemoglobin by Perutz, for which they shared the Nobel Prize in Chemistry in 1962[11]. Both structure determinations relied on the use of isomorphous replacement, with gold and mercury being incorporated into the protein. An indication of the laborious nature of the investigation is clear from the work on myoglobin, which has 2600 atoms in each molecule required the study of 110 crystals and the collection of 250,000 intensities. The structure of myoglobin is shown in Figure 10 in a ribbon form.

Another achievement was the solution of the structure of vitamin B12 by Dorothy Hodgkin (Figure 11), an exceptionally large 'small molecule' with 93 non-hydrogen atoms (Figure 12), for which, together with her solution of the structure of penicillin (18 non-hydrogen atoms; determined in 1945 from manually-calculated 2-dimensional Fourier projections), she was awarded the Nobel Prize in Chemistry 1964[12]. At the time, although crystallography had developed considerably since its inception, solution of the structures of such complex molecules relied on a combination of chemical intuition and sheer perseverance.

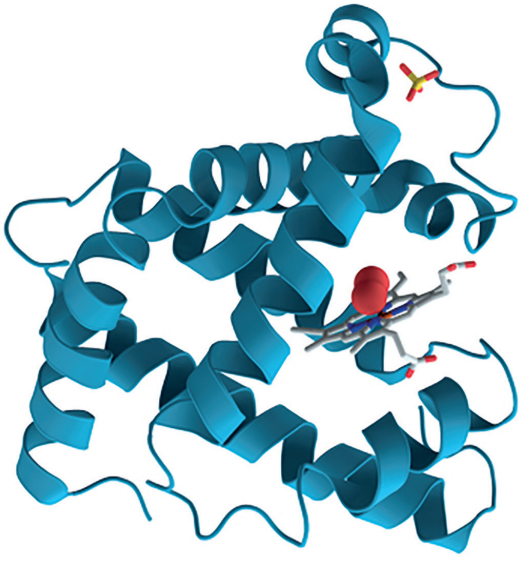


Figure 10 *Structure of myoglobin.*

Figure 11 *Dorothy Crowfoot Hodgkin.*

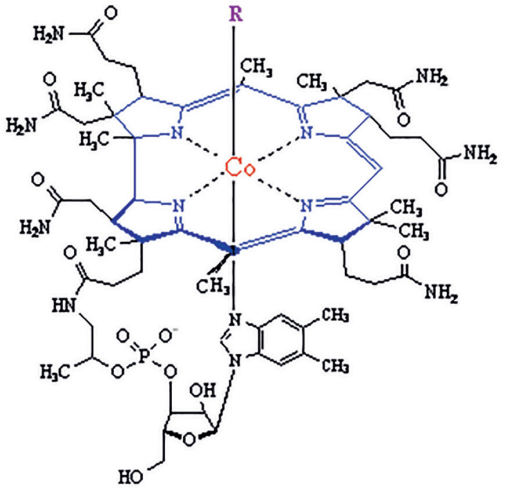


Figure 12 *Structure of Vitamin B12*[12].

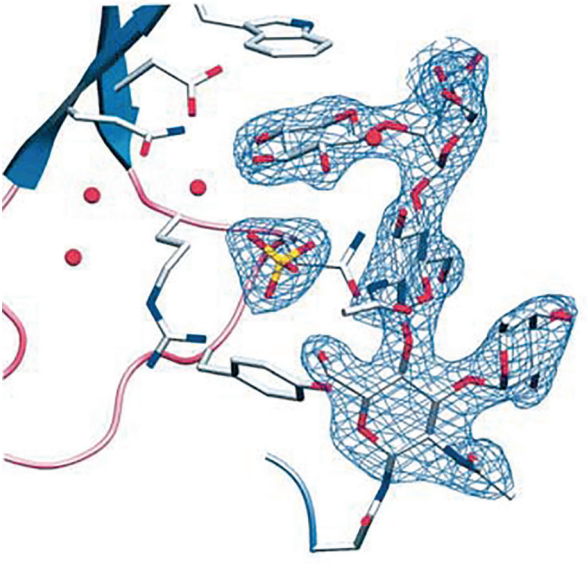


Figure 13 *The SNAII N-terminal xylose-binding site where a terminal xylose unit in the glycosylating chain at Asn-63 binds. Reproduced from ref. 13.*

X-ray crystallography is now used routinely to examine the interaction between a protein target and a pharmaceutical drug. An account of the details of the technique is given by Palmer and Niwa[12] from which we take as an example the binding of a xylose unit in the glycosylation chain at Asn-63 to SNAII protein (Figure 13).

4.2 Structure of DNA

This determination has a convoluted, even controversial, story behind it and it produced probably the most famous XRD photograph of the 20th century (Figure 14).

The key photograph, known as 'Photo 51', was taken by a PhD student, Raymond Gosling in 1952; this exhibits diffraction from an oriented fibre rather than from a single crystal or crystalline powder. Gosling was supervised by Rosalind Franklin at King's College, London. James Watson was shown the photograph by Maurice Wilkins without the knowledge or approval of Franklin. Watson, together with his colleague, Francis Crick, was able to use the features of 'Photo 51' to define the principal structural characteristics of DNA as consisting of a double helix[14]. The outside of the helix has a backbone of alternating deoxyribose and phosphate molecules, while within there are hydrogen-bonded base pairs

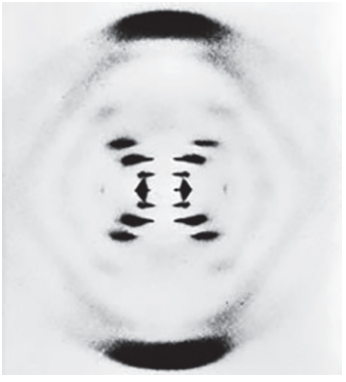


Figure 14 *X-ray pattern from DNA.*

purine adenine (A) always with pyrimidine thymine (T)

pyrimidine cytosine (C) always with purine guanine (G).

The sequence of the sets of pairs delivers the genetic code. For this insight, Watson and Crick were awarded the Nobel Prize in Physiology or Medicine in 1962; a co-awardee was Maurice Wilkins who had determined the essential helical structure of DNA while Watson and Crick had elaborated the internal structure of the double helix and the crucial role of the base pairs[15]. Although Franklin had predeceased the year of the award by four years, there has been a lively, even acrimonious, debate since then concerning the recognition of her role in the discovery.

4.3 Structure of ribosome

A ribosome is a biochemical entity consisting of RNA and associated proteins and is found in large numbers in the cytoplasm of living cells. Its function is to bind messenger RNA (mRNA) and transfer RNA (tRNA) to synthesise polypeptides and proteins. Each ribosome is composed of a large and a small subunit. The subunits are made up of ribosomal (rRNA) molecules, constructed from nucleotides and proteins. The Nobel Prize in Chemistry in 2009 was awarded jointly to Venkataraman Ramakrishnan, Thomas Steitz and Ada Yonath for 'studies of the structure and function of the ribosome'[16]. Yonath produced well-organised crystals with millions of ribosomes assembled into regular patterns and Steitz managed to solve the phase problem to give the first crystal structure of the large subunit; this was improved upon by each of the Laureates to produce a high-resolution structure, enabling the positions of single atoms to be identified. Images of the 70S ribosome are shown in Figure 15[17].

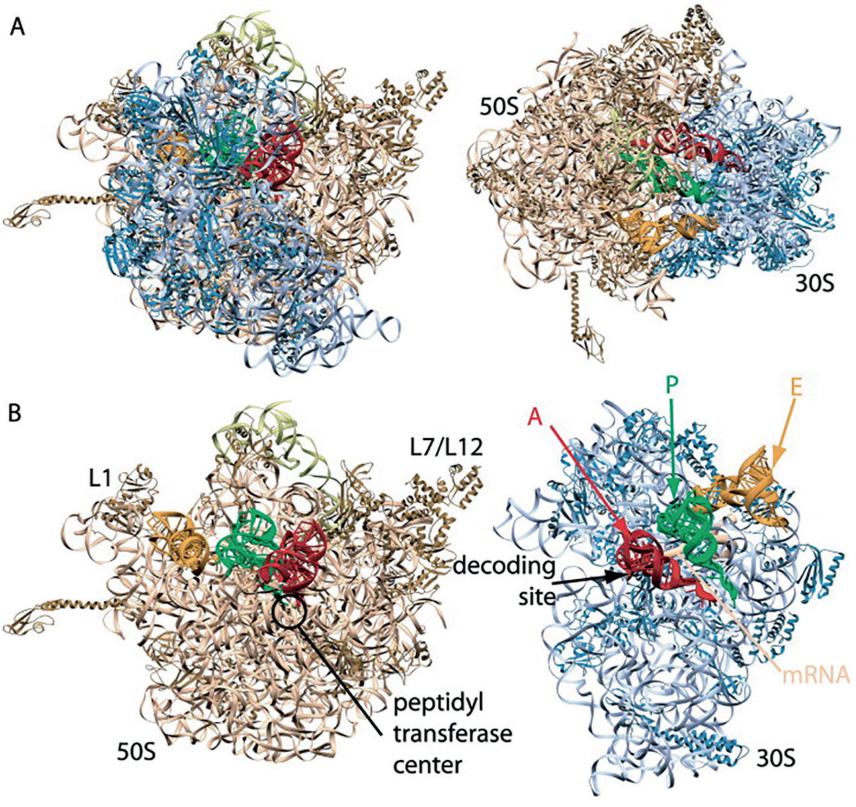


Figure 15 *Crystal structure of the ribosome (A) two views of the 70S ribosome with mRNA and tRNA with the 'top' view on the left and the view from the 30S side on the right and (B) exploded view of the 50S (left) and 30S (right) in the 70S ribosome, showing the locations of the A-, P- and E-sites (from ref. 17).*

4.4 Synthetic polymers

Although for a while it was considered that synthetic polymers were unlikely to display such regularity of structure as to enable diffraction of X-rays, it was found in the 1920s and 1930s that, on stretching, the diffraction pattern of rubber changed from that of an amorphous material to one of crystalline nature[18] as exemplified in Figure 16[19]. The development of stereoregular polymers by Karl Ziegler and Giulio Natta[20] meant that X-ray study could yield valuable information concerning their internal structure. The synthesis of such polymers depended on the use of organometallic catalysts to generate a growing polymer chain under steric control of the catalytic centre.

4.5 Supramolecular structures

There has been a phenomenal growth in work on the devising of complex molecular structures exhibiting more than one different function within a single molecule, each

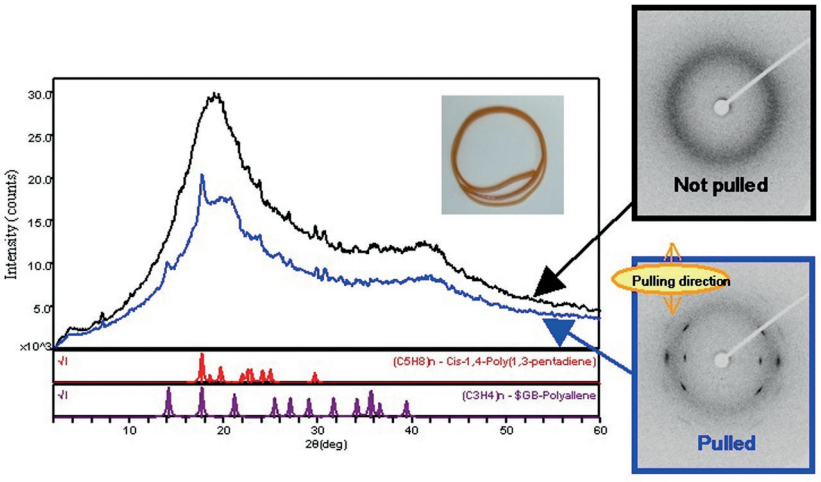


Figure 16 *Diffraction patterns of stretched and unstretched rubber (from ref. 19).*

represented by a subunit; indeed, an entire journal is now devoted to this topic[21]. The importance of this work was recognised in the award of the Nobel Prize in Chemistry in 1987 to Professors Donald J. Cram and Jean-Marie Lehn and Dr. Charles J. Petersen for their development and use of molecules with structure-specific interactions of high selectivity[22]. The structural complexity of the systems means that X-ray crystallography plays a decisive role in establishing both their structure and mode of action. One of the main achievements of this type of system is in providing a means of molecular recognition of biomolecules such as acetylcholine[22]. Many of the recognition systems depend on the construction of large macrocyclic compounds featuring a central 'hole' capable of accommodating a molecule of precise dimensions. The specificity of the macrocyclic ligand can be increased by the presence of donor atoms in the ring, such as O, N, S, *etc.* One example of the encapsulation of an organic molecule is that of *trans*-3,3'-azobenzene dicarboxylate in preference to the *cis*-isomer by the cyclic ligand shown in Figure 17[23].

The last two decades have seen the appearance of some extraordinary structures acting as sensors for specific molecules, such as the anion-specific circular iron(II) helicates depicted in Figure 18[24]. The corresponding crystal structures are shown in Figure 19.

5. The impact of crystallography on chemical bonding theory

From its earliest years, the determinations of crystal structures has had an impact on concepts on the nature of bonding. The ionic, as opposed to covalent, bonding in alkali metal halides was confirmed[25] in 1922 while the structure of diamond confirmed the tetrahedral nature of the carbon atom and the C–C bond length of 1.52 Å[26]. The structure of graphite was discovered in 1942 using X-ray powder diffraction[27] and that of metallic copper in 1922 by W.L. Bragg[28]. The geometry of metal complexes was clarified by the study of the structure of ammonium

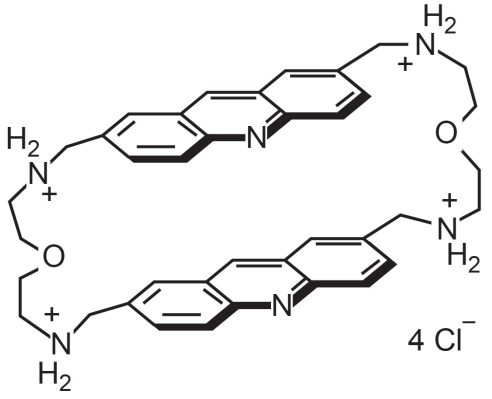
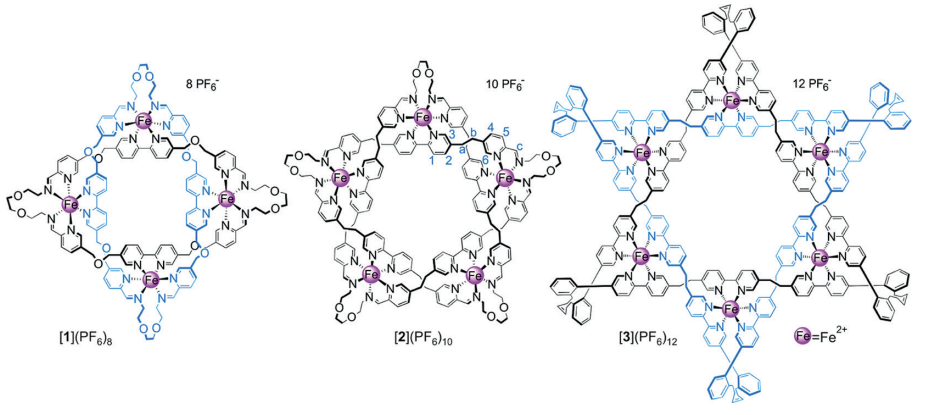


Figure 17 *Structure of encapsulating ligand for cis- and trans-3,3'-azobenzene dicarboxylate[23].*



1 2 3

Figure 18 *Chemical structures of Solomon link (**1**] $(PF_6)_8$, pentafoil knot (**2**) $(PF_6)_9$ and Star of David catenane (**3**) $(Ph_4B)_{11}$. Distances between the centres of the binding pockets and iron ions are depicted by blue double-headed arrows and the diameters of the pockets by black double-headed arrows (from ref. 24).*

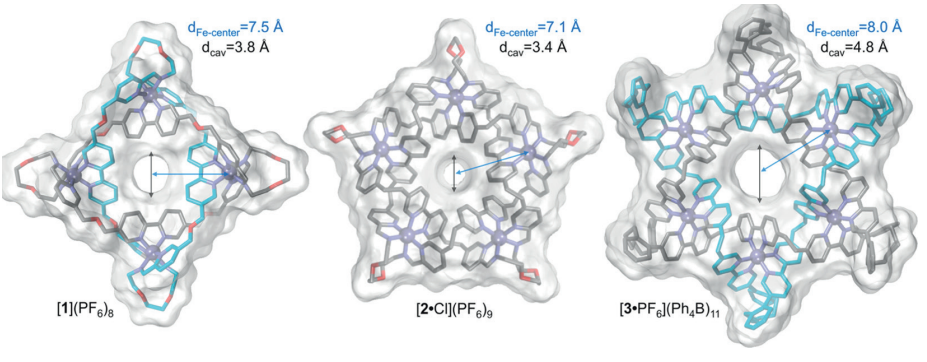


Figure 19 *Crystal structures of Solomon link (**1**) $(PF_6)_8$, pentafoil knot (**2**) $(PF_6)_9$ and Star of David catenane (**3**) $(Ph_4B)_{11}$. Distances between the centres of the binding pockets and iron ions are depicted by blue double-headed arrows and the diameters of the pockets by black double-headed arrows (from ref. 24).*

hexachloroplatinate[29] and those of aromatic molecules such as naphthalene and anthracene confirmed the hexagonal nature of the ring[30]. The structure of ferrocene was an opening chapter in organometallic chemistry[31]. More recently, the structure of C_{60} fullerene has confirmed the existence of a whole new family of carbon isomers[32].

The very low scattering factor for the H atom has caused difficulties in placing H atoms in a structure with accuracy and thus has inhibited development of accurate parameters for hydrogen bonds. However, this was gradually overcome when H is bonded to relatively light atoms (C, O, N) but remains difficult for H bonded to heavy atoms. One notable achievement was the elucidation of the structure of pentaborane by Dulmage and Lipscomb[33] which introduced the concept of three-centre bonding (Figure 20).

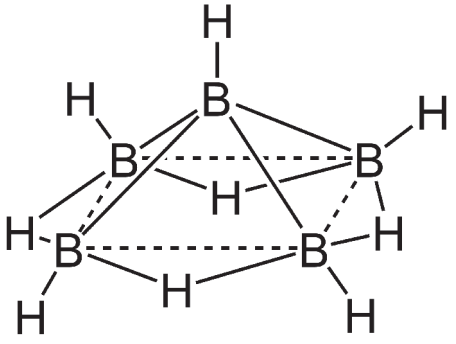


Figure 20 *Structure of pentaborane.*

Lipscomb was awarded the Nobel Prize in Chemistry in 1976 for 'his studies on the structure of boranes illuminating problems of chemical bonding'[34].

5.1 Electron density studies

Fourier syntheses from high-quality diffraction data show the electron density with enough precision to reveal the bonding electrons themselves. A classic study of this type is on oxalic acid dihydrate (Figure 21) which was first reported in 1980 by Stevens and Coppens[35].

The diagram from a more recent study (Figure 21)[36] shows the 'difference electron density', obtained by subtracting the electron density of spherical atoms

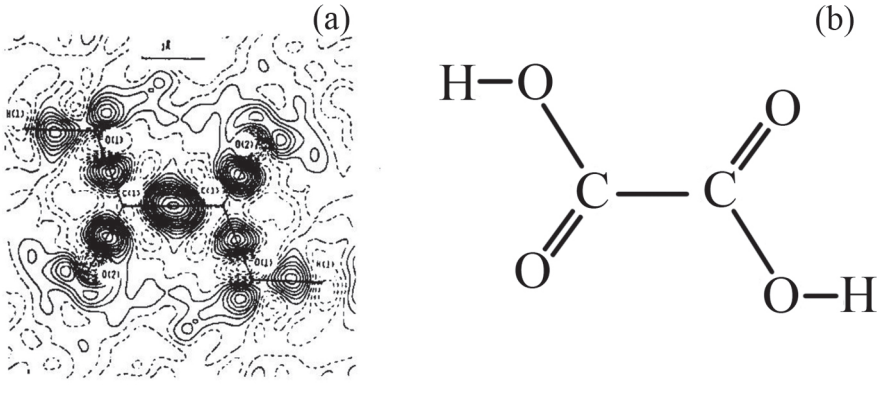


Figure 21 *Difference electron density of oxalic acid dihydrate (a) in the plane of the oxalic acid molecule, oriented as shown in the diagram (b) (from ref. 36).*

from the total electron density. What remains are the bonding electrons and they appear as peaks on the lines joining the atom pairs, exactly in accord with quantum theory.

5.2 Hydrogen bonds

The parameters of hydrogen bonding are best derived from a combination of techniques. The intrinsic problem of weak scattering of X-rays by H atoms is overcome in neutron diffraction, where the scattering of neutrons by H atoms is not proportional to the atomic number (see Section 3.3). An elegant illustration of the application of several techniques is found in the study of cellulose I_α using electron micrography, synchrotron X-ray diffraction and neutron diffraction. The relevant patterns are shown in Figure 22[37].

5.3 Secondary bonding

X-ray crystallography has also revealed unexpected aspects of bonding, for example secondary bonding as defined by Nathaniel Alcock[38]. Structures of compounds of the non-metals have shown bond distances in some instances which are longer than covalent distances but shorter than conventional van der Waals distances. Thus in the structural element

Y–A—X

Y–A will be a normal covalent bond and A—X will have a 'short' interatomic distance. The Y–A–X moiety will be linear, not unlike most hydrogen bonds, but the energies involved are weaker than those found in hydrogen bonds. Figure 23 shows examples of secondary bonds, between 50 and 100 pm longer than the corresponding covalent bonds. The secondary bonds complete the coordination of the central heavy atom, with the examples shown having octahedral, square planar and linear geometries.

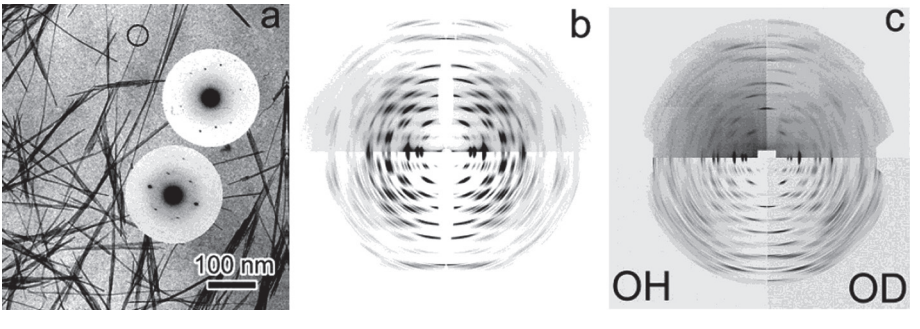


Figure 22 *(a) Electron micrograph of cellulose microcrystals, (b) (top) synchrotron X-ray diffraction data, (bottom) 3D fit of Bragg intensities taking into account fibre texture and (c) (top) neutron fibre diffraction patterns collected from two fibres, one hydrogenated (left quadrant) and one deuterated (right quadrant). The bottom quadrants are 3D fits of the Bragg intensities taking into account fibre structure (from ref. 37).*

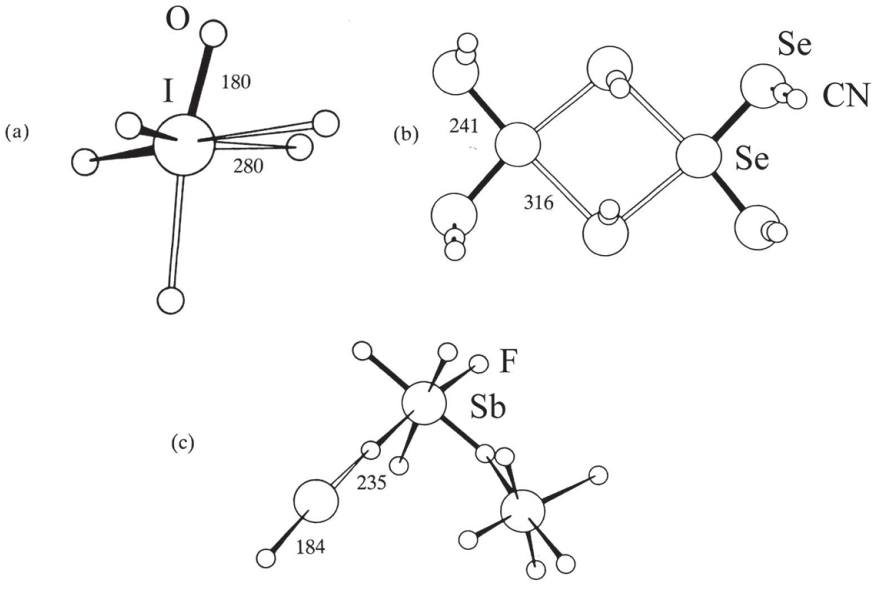


Figure 23 *Secondary bonding (a) octahedral, in the iodate ion in* $[NH_4]^+$ $[IO_3]^-$*, (b) square planar in* $[Se_2(SeCN_6]^{2-}$ *(c) linear in* $[XeF]^+$ $[Sb_2F_{11}]^-$ *(from ref. 38). Dimensions in pm.*

6. Cambridge Structural Database

The Cambridge Crystallographic Data Centre was founded in 1965 by Dr Olga Kennard and began collecting details of the 1,500 previously published organic and organometallic crystal structures. At first, the references were circulated in printed volumes, arranged by type of molecule. However, in 1975 the collected information was converted to electronic form and combined with the actual molecular coordinates that had been determined (claimed to be the first numerical scientific database in the world). Through the co-operation of scientific journals and innumerable individuals, this has continued to grow and the Cambridge Structural Database (CSD) now contains over 800,000 'small molecule' structures (2016). Sophisticated software has enabled searching of the database and display of the results has also been incorporated into the search programs. It can be found at www.ccdc.cam.ac.uk.

More recently, the CSD has been supplemented by the Inorganic Structure Database (ICSD, 180,000 entries) and Crysmet, the database for metals and alloys (160,000 entries). The Protein Data Bank, founded in 1971 (but incorporating data collected earlier) includes the structures of proteins, nucleic acids and other macromolecules, with a total of 120,000 entries.

7. References

1. http://www.nobelprize.org/nobel_prizes/physics/laureates/1915/. [Accessed 8th November 2016].
2. http://www.nobelprize.org/nobel_prizes/physics/laureates/1914/. [Accessed 8th November 2016].

3. D.O.W. (1916) *Sci. Prog.,* **11***,* 483.
4. Patterson, A.K. (1935) *Z. Kristallogr., Kristallgeom., Kristallphys., Kristallchem.*. **A90**, 517.
5. Alcock, N.W., Kemp, T.J. and Leciejewicz, J. (1991) *Inorg. Chim. Acta*, **184**, 203.
6. http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1985/
7. Bruni, G., Gozzo, F., Capsoni, D. *et al.* (2011) *J. Pharm. Sci*., **100**, 2321.
8. Debye, P. and Scherrer, P. (1916) *Phys. Z*., **17**, 277.
9. Pan, Q.Q, Guo, P., Duan, J. *et al.* (2012) *Chin. Sci. Bull*., **57**, 3867–3871. Doi: 10.1007/s11434-012-5245-5.
10. Beevers, C.A. and Lipson, H. (1985) *Aust. J. Phys*., **38**, 263.
11. http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1962/. [Accessed 13 January 2017].
12. http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1964/. [Accessed 13 January 2017]
13. Palmer, R.A. and Niwa, H. (2003) *Biochem. Soc. Trans.,* **31**, 973.
14. Watson, J.D. and Crick, F.H.C. (1953) *Nature*, **171**, 737.
15. http://www.nobelprize.org/nobel_prizes/medicine/laureates/1962/. [Accessed 13 January 2017].
16. http://www.nobelprize.org/nobel_prizes/chemistry/laureates/2009/. [Accessed 13 January 2017]
17. Ramakrishnan, V. (2002) *Cell*, **108**, 557.
18. Lotmar, W. and Meyer, K.H. (1936) *Monatsh. Chem*., **69**, 115.
19. http://www.rigaku.com/en/products/xrd/rapid/app009. [Accessed 13 January 2017].
20. http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1963/. [Accessed 13 January 2017].
21. *Supramolecular Chemistry*, publisher Taylor and Francis, started in 1992-3.
22. http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1987/. [Accessed 13 January 2017].
23. Cudic, P, Vigneron, J.-P., Lehn, J.-M. *et al.* (1999) *Eur. J. Org. Chem*., 2479.
24. Ayme, J.-F., Beves, J.E., Campbell, C.J. *et al.* (2015) *J. Am. Chem. Soc*., **137**, 9812.
25. Bragg, W L., James, R.W. and Bosanquet, C.H. (1922) *Philos. Mag. (1798-1977)*, **44**, 433.
26. Bragg, W.H. and Bragg, W.L., (1913) *Proc. R. Soc. A,* **89**, 277.
27. Lipson, H. and Stokes, A.R. (1942) *Proc. Roy. Soc*. **181**, 101.
28. Bragg, W.L. (1914) *Philos. Mag. (1798-1977)*, **28**, 355.
29. Bragg, W. H. (1921) *Proc. Roy. Soc. Lond.,* **34,** 33.
30. Wycoff, R.W.G. and Posnjak, E. (1921) *J. Am. Chem. Soc.,* **43**, 2292.
31. Dunitz, J.D. Orgel, L.E. and Rich, A. (1956) *Acta Cryst.,* **9**, 373.
32. Liu, S., Lu, Y.-J., Kappes, M.M. and Ibers, J.A. (1991) *Science*, **254**, 408.
33. Dulmage, W.J. and Lipscomb, W.N. (1951) *J. Am. Chem. Soc.,* **73***,* 3539.
34. http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1976/
35. Stevens, E.D. and Coppens, P. (1980), *Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem.,* **36**, 1864.
36. Dam, J, Harkema, S. and Feil, D. (1983) *Acta Crystallogr., Sect. B: Struct. Sci.,* **39**, 760.
37. Nishiyama, Y., Sugiyama, J., Chanzy, H. and Langan, P. (2003) *J. Am. Chem. Soc.,* **125**, 14300.
38. Alcock, N. W. (1972) *Adv. Inorg. Chem. Radiochem*., **15**, 1.