ChemComm
FEATURE ARTICLE

X-Ray crystallography in structural chemistry and molecular biology

J. R. Helliwell and M. Helliwell

Chemistry Department, University of Manchester, Manchester, UK M13 9PL

The X-ray crystallography technique has greatly expanded in scope with the development and use of instrumentation at synchrotron sources, particularly in 'large molecule' protein crystallography and now also in 'small molecule' chemical crystallography. A variety of technical themes are described to illustrate this. Structure–function relationships are then discussed, firstly, in thermotropic liquid crystals, with respect to two octahexylphthalocyanines, and secondly, in the molecular recognition of the receptor binding site of the protein concanavalin A for its cognate sugar ligands.

Introduction

X-Ray crystallography is used to determine the three-dimensional structures of chemical compounds and biological macromolecules. The scope of the technique has been transformed by the harnessing of novel instrumentation developments. This has especially involved synchrotron X-ray sources; applications in protein crystallography are described in the book by J. R. Helliwell[1] and in chemical crystallography by Coppens.[2] Fine tuning of the synchrotron X-ray wavelength is exploitable in protein and nucleic acid structure determination (the so-called MAD method). It is also used in the determination of metal atom incorporation and in the identification of metal atoms of similar atomic number. Such approaches use the anomalous dispersion in atomic X-ray scattering factors. Higher intensity, short wavelength synchrotron X-ray beams, as well as very efficient area detectors and cryo-techniques, are exploited in reaching true atomic resolution in protein crystallography. Likewise, these approaches are applicable in chemical crystallography to obtain such data but from much smaller crystals than hitherto. The precision and accuracy of these biological and chemical structures is therefore improved. Molecular functions usually involve conformational changes. The structural changes involved can be studied *via* the determination of a series of crystal structures. Also, fast data collection techniques can allow the recording of sequences of diffraction patterns with time (as in the Laue method) thus opening up the field of time-resolved protein crystallography. Applications include studies of enzymatic reactions and photostimulable proteins in crystals. The X-ray diffraction data acquisition, time-slicing, techniques in biological studies are similar, in fact, in temperature-resolved perturbation studies of chemical structures. Moreover, the comparison of a sequence of structures of a compound, taken at different temperatures, can reveal conformational variations related to function. Such a molecular system, involving thermotropic liquid crystals, is described in this article. Finally, as a second detailed case study, a description is given of the recognition and binding of sugar molecules at a protein receptor binding site, that of the lectin protein concanavalin A. These structural themes illustrate the power of the X-ray crystallography method today.

Technical Themes

Outline of data collection problems to be addressed

In protein crystallography the major problems to be faced[3] include the inherent weakness of individual reflections, the voluminous amount of data to be collected, and sample radiation damage. In the last 15 years or so these technical difficulties have essentially been solved through a combination of high-powered synchrotron X-ray sources, position sensitive area detectors and cryo-cooling of the sample to liquid nitrogen temperatures. Structure determination of a protein must also deal with the phase problem. This is traditionally solved by the method of multiple isomorphous replacement whereby data sets must be collected for the native unmodified protein and then for at least two different isomorphous heavy-metal derivative crystals. This means that the intrinsic problems listed above must be faced several times over. The phase problem can now be approached by collection of data sets from one crystal at multiple wavelengths around an X-ray absorption edge of a metal atom bound in a protein. Time-resolved studies can also now be considered since rapid data collection of X-ray crystallographic data, even down to picosecond exposures, opens up the prospect of dynamical experiments.[4,5] The recording of 'time-sliced' data sets represents a major data collection task to be organised, obviously under computer control.

In chemical crystallography data collection several key problems and developments still need to be addressed for the further expansion of the method. These include small crystals,[6–8] element selective atom identification and perturbation crystallography (*e.g.* 'temperature resolved').[2] Many of the problems to be tackled in chemical crystallography are then analogous to those in protein crystallography. However, sample radiation damage and the phase problem are rarely difficulties in chemical crystallography today.

Instrumentation: sources, detectors and cryo-techniques in X-ray crystallography

X-Rays generated at synchrotron sources possess special characteristics and advantages for crystallography. Synchrotron X-ray beams have a high intensity, are well collimated and are polychromatic. They are also polarised and have a pulsed time structure, useful for special applications involving dichroism[9] and time-resolved structural studies,[4,5] respectively. The most widespread interests so far reside with the use of the brilliance of the beam (*i.e.* the flux emitted into a given solid angle of divergence from the emitting synchrotron electron beam source area) to overcome weak scattering by a sample, and also being able to selectively tune to a desired X-ray wavelength. The focused X-ray beam should match typical crystal sizes, which are $\leqslant 1$ mm and the beam angular cross fire should match the crystal mosaicity $> 0.001°$.[10] The wavelength range of interest is essentially set by the K absorption edges of the metal atoms that can be bound in a crystal sample for anomalous dispersion methods to be applied.[3,11] Short wavelengths are set by the molybdenum K edge at 0.62 Å and at long wavelength by the calcium K edge at 3.07 Å. These limits encompass many other K and L absorption edges of metal atoms. Of special note though are the K edges of selenium (at 0.98 Å) and bromine (at 0.92 Å), which are used to 'label' proteins and nucleic acids respectively for the MAD (multi-wavelength anomalous dispersion) method.

Position sensitive area detectors are critical to efficiently measure diffraction data sets, which can reach up to 10^6 reflection intensities for protein and virus cases, and even 50000 reflections in chemical crystallography. The detector devices available have evolved through various generations (see for example refs. 1, 12, 13). These have still not yet reached an ideal state. The principal problem is to realise a combination of properties in the one device specification. There are, in addition, special data acquisition problems for time-resolved/perturbation crystallography,[4] and also in MAD involving multiple passes of measurement.[14] For small crystals and for atomic resolution protein crystallography there are still limits set by the detector noise floor and the detector dynamic range of response. Also the sensitivity of a PSD is $< 100\%$ over the full wavelength range of interest, especially at short wavelengths. In addition, image readout times of the devices are still often not quick enough. Nevertheless great strides have been made. Hence, an evolution from photographic film and the single element scintillation counter through to wire chambers, TV detectors, image plates and now charge coupled devices (CCDs) has taken place over a period of 20–30 years. A new device beckons, the pixel detector,[15] which is silicon based (like CCDs), but it should combine the best features, and avoid the problems, of previous devices.

Cryo-cooling of crystal samples is also important. This usually involves liquid-nitrogen temperatures but also includes, in special cases, liquid-helium temperatures. At liquid-nitrogen temperature in chemical crystallography atomic mobility is reduced and the diffraction intensities at high resolution are stronger than if measured with the crystal at room temperature. In protein crystallography the major benefit has been that radiation damage to the crystal has largely been overcome.[16] Cryo-methods have reduced the number of crystals needed to make a complete data set (previously a single virus diffraction data set could otherwise be made up of in excess of 100 crystals!).[17] Moreover cryo-methods have extended the diffraction resolution limit. Some degree of optimisation remains to be realised since the mosaicity of the crystal increases to *ca.* 0.25° on freezing.[18]

Problems of scattering efficiency of crystals and technical solutions

The overall scattering efficiency of a crystal can be calculated according to eqn. (1)

$$\text{Scattering efficiency} = \left(\sum_{i=1}^{K} n_i f_i^2 \right) \frac{V_x}{V_o^2} (\text{e}^2\ \text{Å}^{-3}) \quad (1)$$

where f_i is the atomic number of the ith atom type, n_i is the number of atoms of that type in the cell out of a total of K types, V_x is the sample volume, and V_o is the unit-cell volume.[6] Using this formula crystals of various compositions and volumes can be compared (see Table 1 and the following examples). Generally, a well defined crystal structure of a small molecule requires 50–100 observed reflections for each refined atom. Crystals of 1,4,8,11,15,18,22,25-octahexylphthalocyanine[19] were not sufficiently strongly diffracting using a Mo-Kα sealed-tube source, to allow structural solution. Instead, X-rays from the much higher intensity Cu-Kα rotating anode source were required for full crystal structure analysis. Acceptable numbers of reflections per atom may be obtained using the latter, home laboratory source, on samples with scattering efficiencies down to about 200×10^{12} e^2 Å^{-3}, as for SAPO-43.[20] However, in the room-temperature study of (1,4,8,11,15,18,22,25-octahexylphthalocyanato)nickel (see crystal 1 of Table 1)[21] only 42 reflections per atom were obtainable with this source, although the crystal scattering efficiency was similar to that of SAPO-43. This was because the scattering efficiency formula [eqn. (1)] takes no account of the mobility of the atoms. Thus, whilst crystals of aluminophosphates such as SAPO-43 are generally well ordered, the room-temperature structure determination of (1,4,8,11,15,18,22,25-octahexylphthalocyanato)nickel indicated that atoms of the hexyl groups showed high thermal motion/disorder and this led to very weak diffraction beyond 1.2 Å resolution. In order to improve the reflection to atom ratio, data collection using synchrotron radiation from a multipole wiggler at the Cornell High Energy Synchrotron Source (CHESS), USA, with a CCD detector and cryo-cooling to liquid-nitrogen temperatures was carried out. These data (Fig. 1) allowed full structure solution to high precision and a much improved reflection to atom ratio of 92, even though a somewhat smaller crystal was used (crystal 2 in Table 1).[21] These examples illustrate the new instrumentation capabilities that are now available for the study of weakly diffracting small molecule crystals. The structure–function aspects of these highly substituted phthalocyanines as thermotropic liquid crystals will be described later.

Crystals of protein samples are generally weakly diffracting because of large unit-cell volumes, coupled with high thermal motion and large quantities of disordered water in the crystal. For instance, diffraction was only seen to a resolution of 1.6 Å for the protein concanavalin A of molecular mass 25 kDa, yielding only 17 reflections per atom, where data collection was carried out at room temperature with an image plate area detector and synchrotron radiation from a bending magnet at the DORIS synchrotron source in Hamburg.[22] In such cases then, refinement techniques at atomic resolution used in small molecule crystallography cannot be applied. Instead, restraints on bond lengths and angles, together with isotropic thermal motion refinement are necessary. However, data collection for concanavalin A at liquid-nitrogen temperatures using synchrotron radiation from the multipole wiggler at CHESS, and a CCD detector,[12] allowed observation of data to 0.94 Å resolution. This gave a reflection/atom ratio of 67, a value which would be acceptable for a small molecule sample but now for a structure of 1800 atoms![14,23] Concanavalin A is the largest protein

Table 1 Scattering efficiencies of crystals compared

Sample	Sample volume V_x/μm^3	Overall crystal scattering efficiency /10^{12} e^2 Å^{-3}	Reflections per atom
1,4,8,11,15,18,22,25-octahexylphthalocyanine	16×10^5	1723	71
NiAPO	25×10^4	787	103
CrAPO-14	12×10^4	829	121
SAPO-43	75×10^3	208	91
(1,4,8,11,15,18,22,25-octahexylphthalocyanato)nickel (crystal 1)	12×10^5	184	42
(1,4,8,11,15,18,22,25-octahexylphthalocyanato)nickel (crystal 2)	15×10^4	25	94
Concanavalin A	80×10^5	25	67[a]

[a] 0.94 Å resolution data collection based on CCD + CHESS SR multipole wiggler + cryo-cooling of the protein crystal. By contrast with IP + SR bending magnet + room temperature of the same protein 1.6 Å resolution only could be realised. This arises from use of the CCD being better at measuring weaker signals against X-ray background than an IP, freezing of the protein crystal to protect it from prolonged irradiation and the extremely strong X-ray beam from the multipole wiggler at CHESS to allow the data to be recorded without unduly long exposure times. Based on refs. 14 and 23.

studied at atomic resolution to date. There are other examples (*e.g.* crambin, M_W 3000;[24] rubredoxin, M_W 6000;[25] trypsin, M_W 20 000).[26] There are prospects for extending to even higher molecular mass structures than concanavalin A. However, there will be an upper limit to be expected to the atomic resolution study of proteins in terms of molecular mass (perhaps 50 kDa). Multimacromolecular complexes and viruses are essentially certain to be excluded since they scatter so weakly. Nevertheless, these structures can be studied to a resolution of *ca.* 3 Å, *i.e.* enough to see molecular detail, if not individual atoms. Molecular masses of 10^7 KDa of viruses have been solved by X-ray crystallography. The spherical virus SV40, of 500 Å diameter has been solved at 3.8 Å resolution using synchrotron data collected at CHESS and at the Daresbury Synchrotron Radiation Source (SRS) UK.[27] This is one of the largest structures ever to be solved.

Structure determination by anomalous dispersion methods

When a crystal structure contains an anomalous scatterer it is possible to consider measuring data sets at two different wavelengths to solve the crystallographic phase problem.[3,28–30] Other approaches involve a least-squares determination.[11,31–33] The electron density [eqn. (2)] requires the knowledge of the

$$\rho(x,y,z) = \frac{1}{V}\sum_{hkl} |F_{hkl}|\, e^{i\alpha_{hkl}} e^{-2\pi i(hx+ky+lz)} \qquad (2)$$

phase angle α_{hkl} for each *hkl* reflection but cannot in general be measured directly by a detector.

If one considers the Fourier inverse of this equation [{*i.e.* the structure factor equation [eqn. (3)]} then if the *N*th atom is an

$$F_{hkl} = \sum_{j=1}^{N} f_j e^{2\pi i(hx_j + ky_j + lz_j)} \qquad (3)$$

anomalous scatterer, its scattering factor is complex [eqn. (4)].

$$f_N = f + {}^{\lambda}f' + i^{\lambda}f'' \qquad (4)$$

By measurement of F_{hkl} and $F_{\bar{h}\bar{k}\bar{l}}$ at one wavelength (λ) on the short wavelength side of the absorption edge of the anomalous scatterer (to maximise f'') and F_{hkl} (or $F_{\bar{h}\bar{k}\bar{l}}$) at the wavelength (λ′) right on the absorption edge to make f' most negative, then in principle the changes in intensity of a reflection, *hkl*, are sufficient (with the crystal space group symmetry information) to determine the anomalous scatterer positions, and then to solve for the reflection phase angle, α_{hkl}. Experimentally this has been a challenge to the X-ray crystallographer because the instrumentation needs to tune onto the absorption edge with sufficient wavelength precision and with a small enough wavelength spread ($\delta\lambda/\lambda$). Moreover the diffraction pattern has to be measured very accurately (at the 1% accuracy level in intensity). In spite of these stringent requirements, the use of these MAD techniques for phase determination proves to be practicable. Moreover, the expression of proteins, by use of bacteria grown on a selenomethionine nutrient[11] leads to replacement of sulfur by selenium. Also, the use of metal derivatives (with isomorphism relaxed compared with the traditional method) is possible. Likewise nucleotides can be brominated at a base and the MAD method applied. As an example of the latter,[30] Fig. 2 shows the electron density map of the oligonucleotide d(CGCG^{Br}CG), calculated with phases estimated from diffraction data measured at two very closed spaced wavelengths (only 0.0007 Å apart), namely λ = 0.9185 Å and λ′ = 0.9192 Å using station 9.5 at Daresbury.[34,35]

The determination of transition-metal incorporation in metal substituted aluminophosphates (MeAPOs and MeSAPOs) can also benefit from such techniques. Whilst, for example, in two cases, CoAPO-21[36] and CrAPO-14[20] it was possible to determine the site of incorporation of the transition metal by refinement of the occupancy, due to the correlation of occupancy with the B factor, such methods have limited application. In addition, they are not able to specifically identify metal atoms, necessary particularly when two or more metal

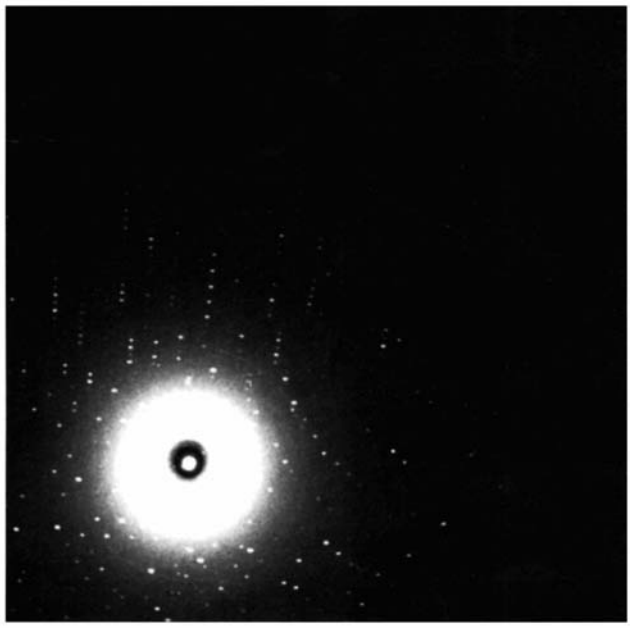

Fig. 1 Synchrotron X-ray diffraction pattern from a very weakly diffracting crystal of (1,4,8,11,15,18,22,25-octahexylphthalocyaninato)nickel (crystal 2 in Table 1) recorded at CHESS (Cornell) Station F2 (multipole wiggler) with a CCD detector. The crystal was cryo-cooled. Wavelength 0.5 Å, angular oscillation range 2°. Exposure time for this image was 90 s. The total number of these exposures needed for the data set was 100. Top edge of the diffraction pattern, where the spots are still just visible, is at 0.7 Å resolution. Based on ref. 21.

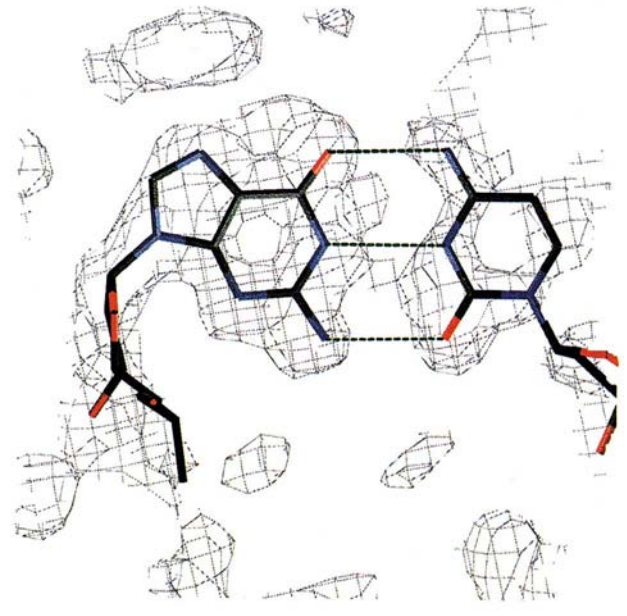

Fig. 2 Crystal structure analysis of a brominated oligonucleotide d(CGCG^{Br}CG) *via* the multiple wavelength anomalous dispersion (MAD) method. Electron density map of a portion of the unit cell calculated from X-ray diffraction data measured on the Daresbury SRS station 9.5[34,35] at two closely spaced wavelengths (λ = 0.9185, 0.9192 Å) exploiting the Br-K absorption edge. Based on ref. 30. It shows that portion of the molecule where there is a Watson–Crick base pair hydrogen-bonding scheme between guanine and cytosine of the oligonucleotide.

atoms, close in atomic number, are present. However, using the anomalous dispersion effects described in eqn. (4) in the previous section, it is possible to locate a metal atom(s). In a centrosymmetric crystal structure the use of two wavelengths (at the absorption edge and remote from the edge) can be used to generate a change in f'. One such example using Cu-Kα and Mo-Kα was the study of NiAPO.[37] The difference in f' for nickel between these two wavelengths is 3.3e$^-$, and calculation of an f' difference-Fourier electron-density map revealed the position of the fully occupied nickel atom. In order to enhance this effect, a further data set was collected close to the f' minimum using synchrotron radiation of 1.4863 Å at the National Synchrotron Light Source (NSLS) in Brookhaven, USA, giving an f' difference between the Mo-Kα and synchrotron data of 7.6e$^-$.[38] The signal was therefore more than doubled. Calculation of the f' difference-Fourier electron density map showed enhanced peak height and coordinate precision due to the larger difference in f'. Since in this case the nickel atom is fully occupied at an octahedral site in the framework, this served as a test case to develop tunable wavelength techniques; in particular, the study demonstrated that such experiments can be further optimised by data collection at several wavelengths using the same instrument, as well as requiring position sensitive detectors to allow more rapid data acquisition. To summarise, the point of developing tunable wavelength techniques in chemical crystallography is twofold: first to study cases of partial substitution (low occupancy) and secondly where two metal atoms, close in atomic number, are present. The first example of the latter case was to determine the Mn and Ca sites in the protein pea lectin.[39]

The direct study of structural changes by X-ray crystallography

Time-resolved macromolecular crystallography[4] is a new capability, but has been of interest for many years.[40] It is driven by continuing improvements in SR sources, optics and detectors. Additionally, the initiation and monitoring of reactions in crystals is new and there are also possibilities, *via* protein engineering and molecular biology, to control the rates of reactions to some extent (*e.g.* see ref. 41). The justification of time-resolved methods comes when structural information from 'static' crystallography of one or a series of protein structures still cannot give key details. For example, a key loop in a protein structure may be invisible in an electron density map from the static crystallography because it is disordered (*i.e.* takes up no particular conformation prior to the chemical reaction) but then orders, temporarily, whilst the enzyme binds to the substrate and undergoes its catalytic reaction cycle. There are a number of technical hurdles to be surmounted to study a particular system by time-resolved methods. There has to be information on the reaction rate for each step, firstly in solution and then in the crystal. For the latter it is possible that the packing of molecules in the crystal lattice blocks access to the active site of the protein or that induced conformational changes create disorder of the crystal. If a reaction can proceed in the crystal it is valuable to monitor UV–VIS spectral changes (*e.g.* where a chromophore is involved)[42] and thus correlate with the variation of the X–ray diffraction data over time.[43]

Transient structures may be exceedingly short-lived. Photosensitive proteins fall into this category. The fastest X-ray diffraction exposures possible arise from use of the full polychromatic SR X-ray beam where diffraction reflections are integrated over wavelength (the Laue method)[29,44–46] rather than by rotation of the crystal (as in monochromatic methods). By use of the point-focused white X-ray beam on the high brilliance wiggler/undulator beamline 3 at the European Synchrotron Radiation Facility (ESRF) in Grenoble, Laue exposures from a myoglobin protein crystal have been recorded from just one electron bunch orbiting the synchrotron ring.[5] This bunch width is 60 picoseconds! The application of this to study the laser light stimulated photodissociation of carbon monoxide bound to the haem in myoglobin is under way,[5] essentially probing then the sub-nanosecond behaviour of this protein structure complex. The application of Laue diffraction to study enzyme reactions in a crystal has stimulated a strong interest with examples including the protein p. 21 in guanosine triphosphate (GTP) hydrolysis[47,48] and isocitrate dehydrogenase[49] as well as photostimulable systems like CO myoglobin.[5] The capability to record accurate protein X-ray diffraction data at the synchrotron with time resolutions from the tens of minutes timescale for monochromatic data and into the sub-nanosecond region for Laue diffraction data has thus been realised, compared to the weeks or months to collect data sets some 15–20 years ago, and thus indeed opening up dynamical studies.[3]

Structural Themes

The structures of thermotropic liquid crystals: two octahexylphthalocyanines compared

In recent years, it has been discovered that some highly substituted octaalkylphthalocyanines are thermotropic liquid crystals. That is, at elevated temperatures, they undergo a phase transition to a discotic columnar mesophase where aromatic cores are formed into columns, but within these columns, individual molecules can adopt any rotational conformation.

The crystal structures of unsubstituted phthalocyanines were amongst the first to be solved directly in the 1930s,[50] with and without metal cations which can be accommodated at the centre of the aromatic cores with replacement of two protons. Unsubstituted phthalocyanines pack into the characteristic herringbone formation where the phthalocyanine planes lie at a distance of approximately 3.4 Å, with their centres offset so that the π cloud of one molecule lies over the cavity of the next, thus maximising the π-stacking interaction. In metallated phthalocyanines, the positive charge of the metal ion has the effect of increasing this attractive force. In the other dimension, the molecules are perpendicular to one another again maximising π–π interaction.

The observation of thermotropic liquid-crystal behaviour[51] in the highly substituted homologues meant that X-ray crystallographic analysis was crucial in determining the modes of packing in these crystals, in order to give insight into the transition to the liquid-crystal phase. Spectroscopic evidence had suggested that a number of modes of packing existed for different homologues. Unfortunately, however, crystals of these materials tend to be fine and very weakly diffracting needles.

The first crystal structure to be determined in this series was of 1,4,8,11,15,18,22,25-octahexylphthalocyanine [Fig. 3(*a*)].[19] Crystals of this compound were weakly diffracting plates, but structure determination was still possible using data collected from a Cu-Kα home-laboratory rotating anode source. Fig. 3(*a*) shows the packing of the molecules into tilted columnar stacks, side by side in two dimensions. Six of the hexyl groups on each molecule lie approximately in the plane of the aromatic core whilst the other two make an angle of approximately 90° to this plane, lying above and below, and acting so as to space the planes to a minimum distance of 8.5 Å. Thus in this case, there can be very little π–π stacking interaction. It has been suggested that raising the temperature would cause the hexyl groups to become mobile,[51] allowing the distance between the aromatic cores to reduce to about 4.4 Å and the discotic columnar mesophase to form at 432 K.

Spectroscopic measurements had indicated that packing in the nickel substituted 1,4,8,11,15,18,22,25–octahexylphthalocyanine was not the same. The crystal structures of this compound [Fig. 3(*b*)–(*e*)] were therefore determined recently both at room temperature (295 K) and at low temperature (103 K). Crystals of this compound were also very weakly diffracting needles. However, data collection using a Cu-Kα rotating anode

diffractometer at 295 K, allowed the structure to be determined, although due to the weak diffraction beyond 1.2 Å resolution (arising from high thermal motion in the hexyl groups), a low number of observations was obtained and the precision of the analysis was rather poor. The space group symmetry was *C*2/*c* and the asymmetric unit consisted of two crystallographically nonequivalent half molecules, the other halves being generated by the operations of an inversion centre and a twofold axis, respectively.

Owing to the low precision of that analysis, a new set of data was collected using X-rays from a multipole wiggler at CHESS to vastly increase the beam intensity, a wavelength of 0.5 Å to compress the diffraction pattern, thereby increasing the recordable resolution (as well as reducing absorption errors) and cryo-cooling to liquid-nitrogen temperatures, to reduce the thermal motion of the molecules, particularly in the hexyl groups. The data were recorded using a CCD detector.[12] Using this regime, it was possible to collect data to 0.7 Å resolution, vastly increasing the data/parameter ratio and the overall precision of the resulting crystal structure. Interestingly it was found that a reduction of space group symmetry had occurred on lowering the temperature, to the space group $P2_1/n$, although the unit-cell dimensions at 295 and 103 K were very similar. At 103 K, the asymmetric unit consisted of two complete molecules. The temperature at which the space group transition occurred was subsequently determined at the SRS by recording, on a CCD[52] and an IP, a sequence of X-ray diffraction patterns, *i.e.* akin to time-resolved data collection. The transition temperature was found to lie between 195 and 205 K.

The overall structures in each space group were similar [Fig. 3(*d*), (*e*)], with the two nonequivalent molecules interleaving one another in tilted stacks. In each case, the lobes of the phthalocyanine cores were staggered allowing a close approach of the almost parallel cores to a minimum distance of 3.3 Å [Fig. 3(*b*), (*c*)]. This indicated that there was considerable π-stacking interaction between the cores, increased further since the positively charged nickel ions were off-set so as to lie above or below the π cloud of the next molecule. The main difference between the crystal structures determined at 103 and 295 K was seen in the degree of order in the hexyl groups. At 103 K, the hexyl groups were for the most part very well ordered. In contrast at 295 K, these groups show high thermal motion/disorder as well as some small conformational differences [Figs. 3(*d*) and (*e*)]. At elevated temperatures, 413 K and above, the thermal motion in these groups presumably increases still further so that the molecules which are locked rotationally by the hexyl groups in the crystalline phase, can rotate freely relative to one another, thus leading to the liquid-crystal phase. Thus the space group transition, which was found to occur at about 200 K, and the subsequent structural changes, mark stages in the transition from the crystalline phase to the discotic columnar mesophase.[21,51]

Saccharide molecular recognition in the protein concanavalin A

Concanavalin A is a protein of the lectin family. Lectins are defined[53,54] as sugar-binding proteins of non immune origin, capable of specific recognition and reversible binding to carbohydrate moieties of complex carbohydrates. Such a function is important especially in cell to cell cross-linking and is also a feature in cell attachment by viral proteins. The basic structure of the sugar-free protein was determined in the 1970s.[55,56] More recently, crystal structures of the protein and its complexes with sugars[57–59] have revealed the intricacies of many aspects of the protein structure, conformational changes and ligand recognition, and also the ordered water structure.[22,23] On monosaccharide binding some waters are expelled[22,58] whereas in another case an ordered water molecule[60] remains. The monosaccharide binding site geometry relies on an unusual *cis*-peptide, stabilised by a calcium ion, which in turn is stabilised by a transition-metal ion (usually a manganese ion).[61] These then are the outlines of the structural chemistry and molecular biology of this protein, and its function [Figs. 4(*a*)–(*f*)]. Although concanavalin A was first isolated in crystalline form, from the jack bean *Canavalia ensiformis*, by Sumner in 1919,[62] its amino acid sequence of 237 residues was not determined until 1975.[63,64] The chemical sequence was corrected later by the 'gene sequence'[65] and showed 15 amino acid changes. More recently, the gene sequence was redetermined and two residues re-assigned back to the chemical sequence.[66]

The determination of the crystal structure of the sugar-bound form of the protein was a problem because the sugar binding site was blocked in the original sugar-free protein crystal form and soaking in of sugar into the protein crystal was not feasible.

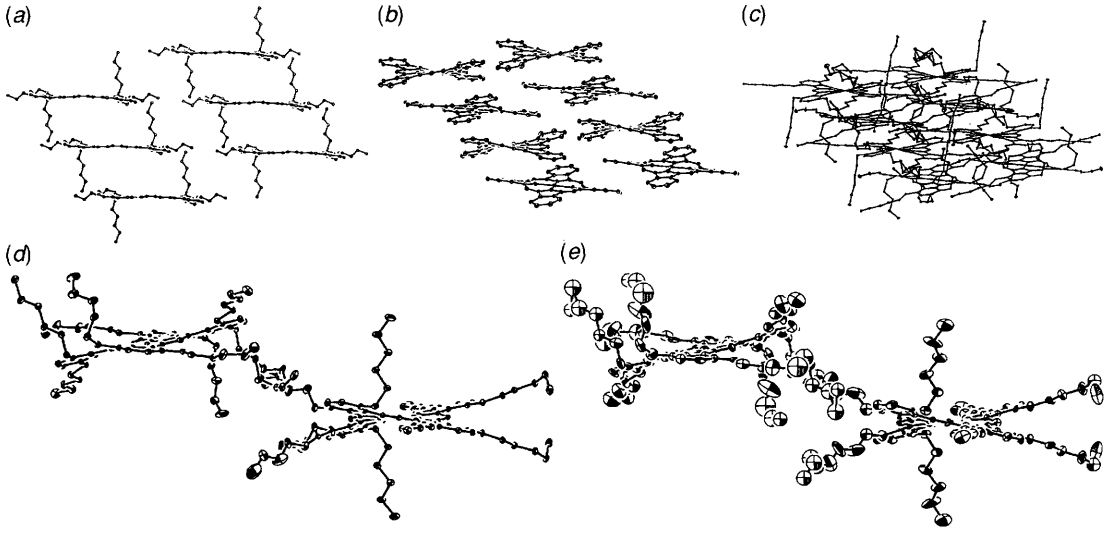

Fig. 3 Towards the liquid-crystal phase of octahexylphthalocyanines. The spacing effect of the hexyl groups in the metal-free 1,4,8,11,15,18,22,25-octahexylphthalocyanine is shown in (*a*) with a layer to layer separation of 8.5 Å. In the nickel substituted form, the packing involves a much closer layer to layer separation to a minimum of 3.3 Å, mediated by the metal; (*b*) shows the tilted stacks, with the hexyl groups omitted for clarity; (*c*) shows the same view including the hexyl groups. In (*d*) and (*e*) are shown the low- (103 K) and room-temperature (295 K) nickel substituted structures where the increased mobility of some of the alkyl chains in the latter is pronounced thus indicating the conformational pathway towards the fully functional liquid-crystal phase at 413 K. Based on refs. 19 and 21.

Hence, new crystal forms of the protein with sugar bound were required. One crystal form, space group $P2_12_12_1$, of the protein with bound methyl α-D-mannopyranoside (one sugar per protein subunit) comprised a whole tetramer in the crystallographic asymmetric unit (in excess of 7000 atoms!). It was solved initially at 2.9 Å resolution and represented the first lectin saccharide structure to be determined.[57] Another crystal form, space group $I2_13$, was for the protein with methyl α-D-glucopyranoside bound (now with one dimer of the tetramer as the unique object).[59] The details of the glucoside binding interactions have been found to be quite similar to the mannoside, in spite of the completely different crystal form. There was one key feature that was different. This was associated with the 2-position hydroxy group which, being equatorial, rather than axial, has fewer binding interactions with the protein receptor site. This then provides the structural explanation of why glucoside binding is weaker than that of mannoside (a factor of three reduction in affinity constant, *ca.* 1 mmol dm^{-3} for glucoside, 0.3 mmol dm^{-3} for mannoside). High-resolution, refined crystal structures of the sugar-free form (at 1.6 Å[22] and now at 0.94 Å[23]), the mannoside bound (at 2.0 Å)[58] and the glucoside bound (at 2.0 Å)[59,67] as well as their detailed comparison finally shows the fine details of the protein ligand interactions. In particular three water molecules in the binding site are displaced by the sugar, having previously occupied the positions of the oxygen atoms of the bound sugar (see front cover). The entropic aspects of this binding process are then quite intricate since both the sugar (presumably also hydrated before binding) and the protein lose their bound waters. Also, the sugar makes a set of hydrogen bonds and van der Waals interactions with the protein ligand atoms. The side-chain of aspartic acid 208 is of particular note, being hydrogen bonded to both O6 and O4 of the sugar, as well as being involved in coordination *via* a water molecule to the calcium ion. Moreover the alanine 207 to aspartic acid 208 peptide link is *cis* and stabilised by the calcium ion. This bond appears very clearly in a *trans* conformation in the metal-free concanavalin A,[61] for which the Asp 208 side-chain then points away from the sugar binding site. Hence, there has evolved an intricate binding sequence. A transition-metal ion with an octahedral arrangement of ligands is required first and then a calcium ion in a seven-coordinate pseudo-octahedral arrangement of ligands. The latter stabilises the unfavourable *cis* peptide, which creates key features of the Asp 208 side-chain orientation to facilitate, finally, the sugar binding. The story has developed further recently. The biological activities of concanavalin A depend on the specific binding of the protein to polysaccharide receptors, in particular a trimannoside core [3,6-di-O-(α-D-mannopyranosyl)-α-D-mannopyranoside]. This has an affinity some 60 times higher than methyl αmannose (MeαMan). This is in contrast to other plant lectins, which show no such enhanced binding whilst their monosaccharide binding is very similar to the structural results found for concanavalin A. The concanavalin A trimannoside core structure has been solved independently by two groups,[60,68] with slightly different crystal forms. Both have a tetramer in the asymmetric unit. The details of the MeαMan structural interactions are identical to the monosaccharide, mannoside, case. However, for the additional core sugar there is a conserved water molecule which is retained so as to specifically orient its hydrogen atoms to hydrogen bond

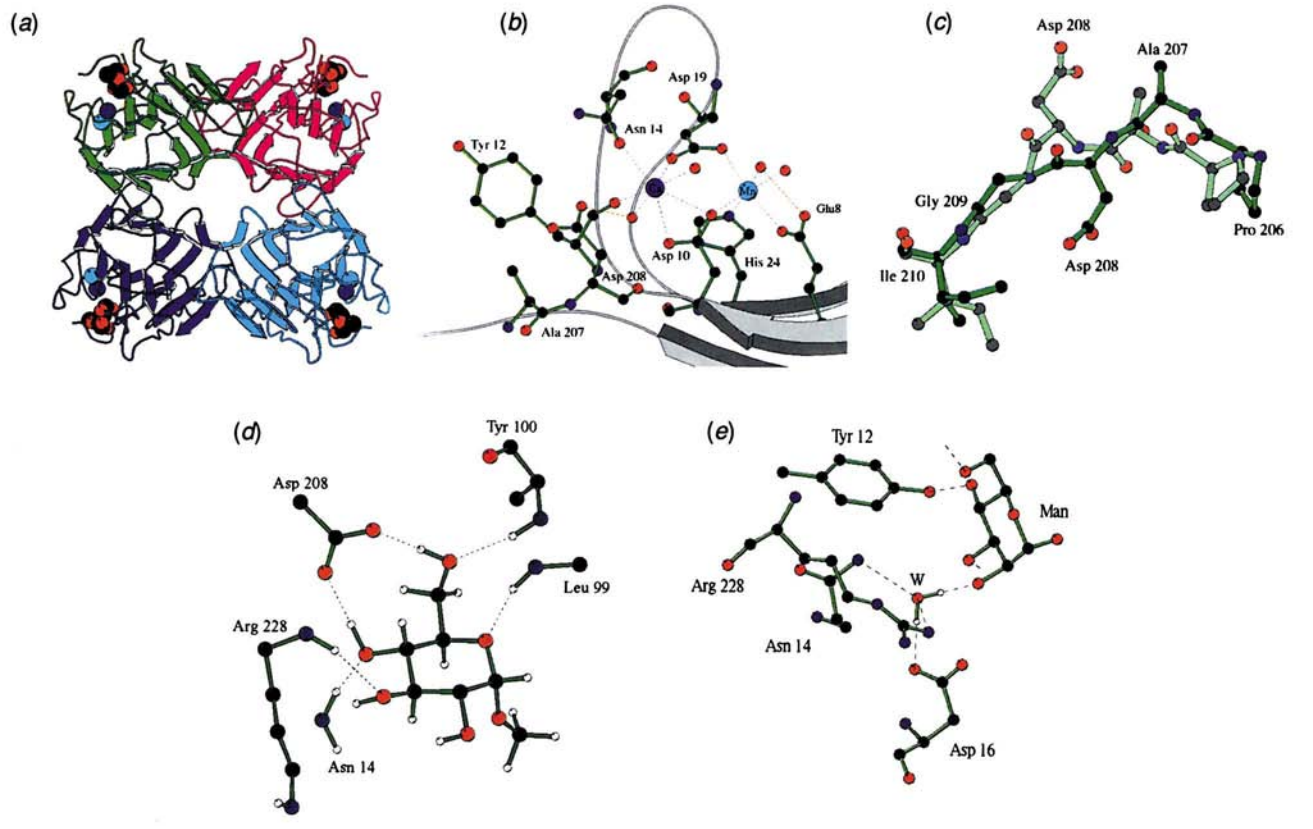

Fig. 4 Concanavalin A crystal structure studies. (*a*) Tetramer of concanavalin A in a ribbon representation showing the extensive 'β-sheet' polypeptide strands (as arrows) with each subunit of the protein colour coded. In addition in each subunit there are shown the essential Mn (light blue) and Ca (dark blue) ions, which are needed to bind a sugar molecule (carbon atoms in black, oxygens in red). Based on ref. 57. (*b*) Close in view of the Mn and Ca binding sites. Note that the Mn ligands form an octahedral coordination and for Ca a pseudo-octahedral (7 ligands) coordination. (*c*) The conversion of the demetallised (open bonds) to metallised (green bonds) structures, and whereby the *cis*-peptide in the latter requires bound Ca. Based on ref. 61. (*d*) The monosaccharide binding site of concanavalin A showing glucoside bound (two OH equatorial); hydrogen bonds are shown dotted. Note the importance of the correct positioning of Asp 208 to facilitate the binding of the sugar. Based on refs. 59 and 67. (*e*) The 'core sugar' in the study of trimannoside binding to concanavalin A.[60] Note the ordered water molecule (W), which assists in the binding of this sugar. This is in contrast to the monosaccharide binding site [Fig. 4(*d*)] where three previously ordered water molecules, see front cover picture, are replaced by sugar oxygens.

to that sugar. The lone pairs of the oxygen are also involved in a hydrogen bond to the protein. The proton positions of this 'structural water' are of course not directly determined from the X-ray analysis but are inferred.

In summary, the molecular recognition details of the monosaccharide to concanavalin A, and also of the trimannoside core, have now been established. These structures now also can serve as a test-bed for the direct calculation of the energetics of binding, experimental values for which are given in the literature from microcalorimetry. It is through such structures and calculations that a quantitative basis exists for the prediction of binding strengths of other sugars and for the impact of altered amino acids, through 'site-specific protein engineering'. These are also fundamental considerations in the design of inhibitors for this protein as a 'drug-design target', or of other proteins.

Concluding remarks

Many technical developments have taken place in the last two decades, especially involving synchrotron radiation, in expanding the capability of the X-ray crystallography technique. The methods and the results span physics to chemistry and biology, and underpin new developments in molecular design applicable to materials science and medicine.

Acknowledgements

We are grateful to the EPSRC, BBSRC, The Wellcome Trust and the EU for funding as well as to the synchrotron radiation facilities in Daresbury (SRS), Cornell (CHESS), Hamburg (DORIS), Brookhaven (NSLS) and Grenoble (ESRF), featured in the studies reported. We are grateful to G. Bradbrook, A. Deacon, T. Gleichmann and M. Peterson for assistance with the various figures. Permission to reproduce Figs. 4(*c*) and 4(*e*) was given by Dr J. Bouckaert of the Vrije Universiteit Brussels and Dr J. H. Naismith of St. Andrews University respectively, to whom we are also very grateful.

Professor John R. Helliwell is Professor of Structural Chemistry at the University of Manchester. He is active in crystallography techniques and the structural studies of lectin proteins and enzymes. He has a B.Sc. First Class Honours and a D.Sc. in Physics from the University of York and a D.Phil. in Molecular Biophysics from the University of Oxford. He is a Fellow of the Institute of Physics and a Fellow of the Royal Society of Chemistry. He is currently Chairman of the MACCHESS Advisory Committee at Cornell University, USA. He has previously served as Chairman of the Daresbury, UK, SRS Beam Allocation Panel for Protein Crystallography, as well as Vice Chairman, and then Chairman, of the European Synchrotron Radiation Facility's Science Advisory Committee in Grenoble, France. He is a Main Editor of the *Journal of Synchrotron Radiation* and a Co-editor of *Acta Crystallographica*.

Dr Madeleine Helliwell is a Research Scientist in the Chemistry Department of the University of Manchester. She is active in chemical crystallography techniques and structural studies of molecular sieves and phthalocyanine thermotropic liquid-crystal structures. She has a B.Sc. First Class Honours in Chemistry (with subsidiary Mathematics) from the University of Reading and a D.Phil. in Synthetic Inorganic Chemistry from the University of Oxford. She is a Member and Chartered Chemist of the Royal Society of Chemistry.

References

1 J. R. Helliwell, *Macromolecular Crystallography with Synchrotron Radiation*, Cambridge University Press, 1992.
2 P. Coppens, *Synchrotron Radiation Crystallography*, Academic Press, 1992.
3 J. R. Helliwell, *Daresbury Laboratory Proceedings DL/SCI/R13*, 1979, 1.
4 D. W. J. Cruickshank, J. R. Helliwell and L. N. Johnson, *Time-Resolved Macromolecular Crystallography*, The Royal Society and Oxford University Press, 1992.
5 D. Bourgeois, T. Ursby, M. Wulff, C. Pradervand, A. Legrand, W. Schildkamp, S. Labouré, V. Srajer, T. Y. Teng, M. Roth and K. Moffat, *J. Synchrotron Rad.*, 1996, **3**, 65.
6 S. J. Andrews, M. Z. Papiz, R. McMeeking, A. J. Blake, B. M. Lowe, K. R. Franklin, J. R. Helliwell and M. M. Harding, *Acta Crystallogr., Sect. B*, 1988, **44**, 73.
7 E. H. Snell, J. Habash, M. Helliwell, J. R. Helliwell, J. Raftery, V. Kaucic and J. W. Campbell, *J. Synchrotron Rad.*, 1995, **2**, 22.
8 M. M. Harding, *Chem. Br.*, 1990, 956.
9 D. H. Templeton and L. Templeton, *J. Synchrotron Rad.*, 1995, **2**, 31.
10 E. H. Snell, S. Weisgerber, J. R. Helliwell, E. Weckert, K. Holzer and K. Schroer, *Acta Crystallogr., Sect. D*, 1995, **51**, 1099.
11 W. Hendrickson, *Science*, 1991, **254**, 51.
12 S. Gruner and S. Ealick, *Structure*, 1995, **3**, 13.
13 N. M. Allinson, *J. Synchrotron Rad.*, 1994, **1**, 54.
14 A. Deacon, J. Habash, S. J. Harrop, J. R. Helliwell, W. N. Hunter, G. A. Leonard, M. Peterson, A. Haedener, A. J. Kalb (Gilboa), N. M. Allinson, C. Castelli, K. Moon, S. McSweeney, A. Gonzalez, A. W. Thompson, S. Ealick, D. M. Szebenyi and R. Walter, *Rev. Sci. Instrum.*, 1995, **66**, 1287.
15 G. Hall, *Q. Rev. Biophys.*, 1995, **28**, 1.
16 H. Hope, *Acta Crystallogr., Sect. B*, 1988, **44**, 22.
17 M. G. Rossman, E. Arnold, J. W. Erickson, E. A. Frankenberger, J. P. Griffith, H.-J. Hecht, J. E. Johnson, G. Kamer, M. Luo, A. G. Mosser, R. R. Rueckert, B. Sherry and G. Vriend, *Nature*, 1985, **317**, 145.
18 E. P. Mitchell and E. F. Garman, *J. Appl. Crystallogr.*, 1994, **27**, 1070.
19 I. Chambrier, M. J. Cook, M. Helliwell and A. K. Powell, *J. Chem. Soc., Chem. Commun.*, 1992, 444.
20 M. Helliwell, V. Kaucic, G. M. J. Cheetham, M. M. Harding, B. M. Kariuki and P. J. Rizkallah, *Acta Crystallogr., Sect. B*, 1993, **49**, 413.
21 M. Helliwell, A. Deacon, K. J. Moon, A. K. Powell and M. J. Cook, submitted.
22 C. Emmerich, J. R. Helliwell, M. Redshaw, J. H. Naismith, S. J. Harrop, J. Raftery, A. J. Kalb (Gilboa), J. Yariv, Z. Dauter and K. S. Wilson, *Acta Crystallogr., Sect. D*, 1994, **50**, 756.
23 A. Deacon, T. Gleichmann, S. J. Harrop, J. R. Helliwell and A. J. Kalb (Gilboa), CHESS Newsletter, 1995, 11–12 and *Rev. Sci. Instrum.*, in the press.
24 M. M. Teeter and H. Hope, *Ann. N.Y. Acad. Sci.*, 1986, 163.
25 G. Sheldrick, Z. Dauter, K. S. Wilson, H. Hope and L. Sieker, *Acta Crystallogr., Sect. D*, 1994, **49**, 18.
26 Th. R. Schneider, K. S. Wilson and F. Parak, Abstract MO70 of the ACA Montreal Meeting, 1995.
27 R. C. Liddington, Y. Yan, J. Moulai, R. Sahli, T. I. Benjamin and S. C. Harrison, *Nature*, 1991, **354**, 278.
28 Y. Okaya and R. Pepinsky, *Phys. Rev.*, 1956, **103**, 1645.
29 J. R. Helliwell, *Rep. Prog. Phys.*, 1984, **47**, 1403.
30 M. R. Peterson, S. J. Harrop, S. M. McSweeney, G. A. Leonard, A. W. Thompson, W. N. Hunter and J. R. Helliwell, *J. Synchrotron Rad.*, 1996, **3**, 24.
31 J. Karle, *Appl. Opt.*, 1967, **6**, 2132.
32 J. Karle, *Int. J. Quantum Chem. Quantum Biol. Symp. 1*, 1980, 357.
33 W. Hendrickson, *Trans. Am. Crystallogr. Assoc.*, 1985, **21**, 11.
34 R. C. Brammer, J. R. Helliwell, W. Lamb, A. Liljas, P. R. Moore, A. W. Thompson and K. Rathbone, *Nucl. Instrum. Methods Phys. Res. Sect. A*, **271**, 678.
35 A. W. Thompson, J. Habash, S. J. Harrop, J. R. Helliwell, C. Nave, P. Atkinson, S. S. Hasnain, I. D. Glover, P. R. Moore, N. Harris, S. Kinder and S. Buffey, *Rev. Sci. Instrum.*, 1992, **63**, 1062.
36 G. M. Cheetham, M. M. Harding, P. J. Rizkallah, V. Kaucic and N. Rajic, *Acta Crystallogr., Sect. C*, 1991, **47**, 1361.
37 M. Helliwell, B. Gallois, B. Kariuki, V. Kaucic and J. R. Helliwell, *Acta Crystallogr., Sect. B*, 1993, **49**, 428.
38 M. Helliwell, J. R. Helliwell, J. C. Hanson, T. Ericsson, A. Kvick, V. Kaucic, C. Frampton and A. Cassetta, *Acta Crystallogr., Sect. B*, 1996, **52**, 479.
39 H. Einspahr, K. Suguna, F. L. Suddath, G. Ellis, J. R. Helliwell and M. Z. Papiz, *Acta Crystallogr., Sect. B*, 1985, **41**, 336.

40 D. C. Phillips, personal communication, 1994.
41 A. C. Niemann, P. K. Matzinger and A. Haedener, *Helv. Chim. Acta*, 1994, **77**, 1791.
42 A. T. Hadfield and J. Hajdu, *J. Appl. Crystallogr.*, 1993, **26**, 839.
43 N. M. Allinson, P. D. Carr, M. Colapietro, M. M. Harding, J. R. Helliwell and A. W. Thompson, *Phase Transitions*, 1992, **39**, 145.
44 K. Moffat, D. Szebenyi and D. H. Bilderback, *Science*, 1984, **223**, 1423.
45 J. R. Helliwell, *J. Mol. Structure*, 1985, **130**, 63.
46 D. W. J. Cruickshank, J. R. Helliwell and K. Moffat, *Acta Crystallogr., Sect. A*, 1987, **43**, 656.
47 I. Schlichting, S. C. Almo, G. Rapp, K. S. Wilson, K. Petratos, A. Lentfer, A. Wittinghofer, W. Kabsch, E. F. Pai, G. A. Petsko and R. S. Goody, *Nature*, 1990, **345**, 309.
48 A. J. Scheidig, A. Sanchez-Llorente, A. Lautwein, E. F. Pai, J. E. T. Corrie, G. P. Reid, A. Wittinghofer and R. S. Goody, *Acta Crystallogr., Sect. D*, 1994, **50**, 512.
49 J. M. Bolduc, D. H. Dyer, W. G. Scott, P. Singer, R. M. Sweet, D. E. Koshland Jr. and B. L. Stoddard, *Science*, 1995, **268**, 1312.
50 J. M. Robertson, *J. Chem. Soc.*, 1936, 1195.
51 M. J. Cook, S. J. Cracknell and K. J. Harrison, *J. Mater. Chem.*, 1991, **1**, 703.
52 K. Moon, N. M. Allinson and J. R. Helliwell, *Nucl. Instrum. Methods, Phys. Res. Sect. A*, 1994, **348**, 631.
53 N. Sharon and H. Lis, *Chem. Br.*, 1990, **26**, 679.
54 I. J. Goldstein, C. M. Reichert and A. M. Misaki, *Ann. N.Y. Acad. Sci.*, 1974, **234**, 283.
55 J. Becker, G. N. Reeke Jr., B. A. Cunningham and G. M. Edelman, *Nature*, 1976, **259**, 406.
56 K. D. Hardman, R. C. Agarwal and M. J. Freiser, *J. Mol. Biol.*, 1982, **157**, 69.
57 Z. Derewenda, J. Yariv, J. R. Helliwell, A. J. Kalb (Gilboa), E. J. Dodson, M. Z. Papiz, T. Wan and J. W. Campbell, *EMBO J.*, 1989, **8**, 2189.
58 J. H. Naismith, C. Emmerich, J. Habash, S. J. Harrop, J. R. Helliwell, W. N. Hunter, J. Raftery, A. J. Kalb (Gilboa) and J. Yariv, *Acta Crystallogr., Sect. D*, 1994, **50**, 847.
59 S. J. Harrop, J. R. Helliwell, T. Wan, A. J. Kalb (Gilboa), L. Tong and J. Yariv, *Acta Crystallogr., Sect. D*, 1996, **52**, 143.
60 J. H. Naismith and R. A. Field, *J. Biol. Chem.*, 1996, **271**, 972.
61 J. Bouckaert, R. Loris, F. Poortmans and L. Wyns, *Proteins, Struct. Funct. Genet.*, 1995, **23**, 510.
62 J. B. Sumner, *J. Biol. Chem.*, 1919, **37**, 137.
63 J. L. Wang, B. A. Cunningham, M. J. Waxdel and G. M. Edelman, *J. Biol. Chem.*, 1975, **250**, 1490.
64 B. A. Cunningham, J. L. Wang, M. J. Waxdel and G. M. Edelman, *J. Biol. Chem.*, 1975, **250**, 1503.
65 D. M. Carrington, A. Auffret and D. E. Hanke, *Nature*, 1985, **313**, 64.
66 W. Min, A. J. Dunn and D. H. Jones, *EMBO J.*, 1992, **11**, 1303.
67 S. J. Harrop, Ph.D. Thesis, University of Manchester, 1992.
68 J. Bouckaert, Ph.D. Thesis, Vrije Universiteit Brussels, 1996.

Received, 21st March 1996; 6/01961G