International Tables for Crystallography (2006). Vol. F, Section 26.1.2, pp. 745–753.

26. A HISTORICAL PERSPECTIVE

26.1. How the structure of lysozyme was actually determined

BY C. C. F. BLAKE, R. H. FENN, L. N. JOHNSON, D. F. KOENIG, G. A. MAIR, A. C. T. NORTH, J. W. H. OLDHAM, D. C. PHILLIPS, R. J. POLJAK, V. R. SARMA AND C. A. VERNON

26.1.1. Introduction

For protein crystallographers, the year 1960 was the spring of hope. The determination of the three-dimensional structure of sperm-whale myoglobin at 2 Å resolution (Kendrew *et al.*, 1960) had shown that such analyses were possible, and the parallel study of horse haemoglobin at 5.5 Å resolution (Perutz *et al.*, 1960) had shown that even low-resolution studies could, under favourable circumstances, reveal important biological information. All seemed set for a dramatic expansion in protein studies.

At the Royal Institution in London, two of us (CCFB and DCP) had used the laboratory-prototype linear diffractometer (Arndt & Phillips, 1961) to extend the myoglobin measurements to 1.4 Å resolution for use in refinement of the structure (Watson *et al.*, 1963), and we had begun a detailed study of irradiation damage in the myoglobin crystals (Blake & Phillips, 1962). Meanwhile, David Green, an early contributor to the haemoglobin work (Green *et al.*, 1954), and ACTN had initiated a study of β-lactoglobulin (Green *et al.*, 1956) and worked together on oxyhaemoglobin before Green went to the Massachusetts Institute of Technology (MIT) in 1959 on leave for a year. At roughly this time, many of the participants in the myoglobin and haemoglobin work at Cambridge went off to other laboratories to initiate or reinforce other studies. Thus, Dick Dickerson went with Larry Steinrauf to the University of Illinois, Urbana, to start a study of the triclinic crystals of hen egg-white lysozyme.

RJP went to MIT from the Argentine as a post-doctoral fellow in 1958 and worked initially with Martin Buerger. In 1959 he transferred to Alex Rich's laboratory and there he soon came into contact with a number of veterans of the myoglobin and haemoglobin work. In addition to David Green were Howard Dintzis, who had discovered a number of the important heavy-atom derivatives of myoglobin (Bluhm *et al.*, 1958) and was now on the staff at MIT, and David Blow, who had first used multiple isomorphous replacement and anomalous scattering to determine haemoglobin phases (Blow, 1958) and was on leave from Cambridge. The influence of these people, combined with lectures by John Kendrew and then by Max Perutz on visits to MIT, soon convinced RJP that working on the three-dimensional structures of proteins was the most challenging and fruitful research that a crystallographer could undertake. Dintzis, in particular, persuaded him that preparing heavy-atom derivatives was no great problem, and Blow urged him to look for commercially available proteins that were known to crystallize. This soon focused his attention also on hen egg-white lysozyme (Fleming, 1922), but in the tetragonal rather than the triclinic crystal form. He quickly learned to grow crystals by the method described by Alderton *et al.* (1945) and then found that precession photographs of crystals soaked in uranyl nitrate showed intensities that differed significantly from those given by the native crystals. Encouraged by these results, he asked Max Perutz whether he could join the Cambridge Laboratory, but Max, having no room in Cambridge, suggested that he write to Sir Lawrence Bragg about going to the Royal Institution. Bragg replied with an offer of a place to work on β-lactoglobulin with David Green, who had by then returned to London. RJP accepted the offer and left for London late in 1960 – after first discussing what was going on at the Royal Institution with ACTN, who had just arrived at MIT for a year's leave with Alex Rich.

Early in 1961, RJP showed Bragg his precession photographs of potential lysozyme derivatives, and Bragg enthusiastically encouraged him to continue the work, at the same time urging DCP to arrange as much support as possible. This was a characteristic response by Bragg, who was well aware that at least two other groups were already working on lysozyme, Dickerson and Steinrauf at Urbana and Pauling and Corey at Cal Tech (Corey *et al.*, 1952): competition with Pauling was a common feature of his career. In describing his reaction to Bragg's encouragement, RJP recalled Metchnikoff's view of Pasteur. 'He transferred his enthusiasm and energy to his colleagues. He never discouraged anyone by the air of scepticism so common among scientists who had attained the height of their success . . . He combined with genius a vibrant soul, a profound goodness of heart.'

26.1.2. Structure analysis at 6 Å resolution

26.1.2.1. *Technical facilities*

In 1961, the Davy Faraday Laboratory was well equipped with X-ray generators. They included both conventional X-ray tubes, operating at 40 kV and 20 mA to produce copper $K\alpha$ radiation, and high-powered rotating-anode tubes that had been built in the laboratory to the design of D. A. G. Broad (patent 1956) under the direction of U. W. Arndt. We had a number of Buerger precession cameras and a Joyce–Loebl scanning densitometer, which had been used in the analysis of myoglobin (Kendrew *et al.*, 1960). In addition, we had a laboratory prototype linear diffractometer (Arndt & Phillips, 1961), which had been made in the laboratory workshop by T. H. Faulkner, and the manually operated three-circle diffractometer that had been used to make some of the measurements in the 6 Å studies of myoglobin (Kendrew *et al.*, 1958) and haemoglobin (Cullis *et al.*, 1961). The diffractometers were used with sealed X-ray tubes, since the rotating anodes were not considered to be reliable or stable enough for this purpose.

At this stage, most of the computations were done by hand, but we did have access to the University of London Ferranti MERCURY computer, usually in the middle of the night. This machine was programmed in MERCURY Autocode. The development of the early computers, their control systems and compilers mentioned in this article have been described by Lavington (1980).

26.1.2.2. *Lysozyme crystallization*

Tetragonal lysozyme crystals were first reported by Abraham & Robinson (1937) and the standard method of preparation was developed by Alderton *et al.* (1945); RJP used this method. Lyophilized lysozyme was obtained commercially and dissolved in distilled water at concentrations ranging from 50 to 100 mg ml^{-1}. To a volume of the lysozyme solution, an equal volume of 10% (*w*/*v*) NaCl in 0.1 *M* sodium acetate (pH 4.7) was added. About 1 to 2 ml aliquots of this mixture were pipetted into glass vials and tightly capped. Large crystals, frequently with

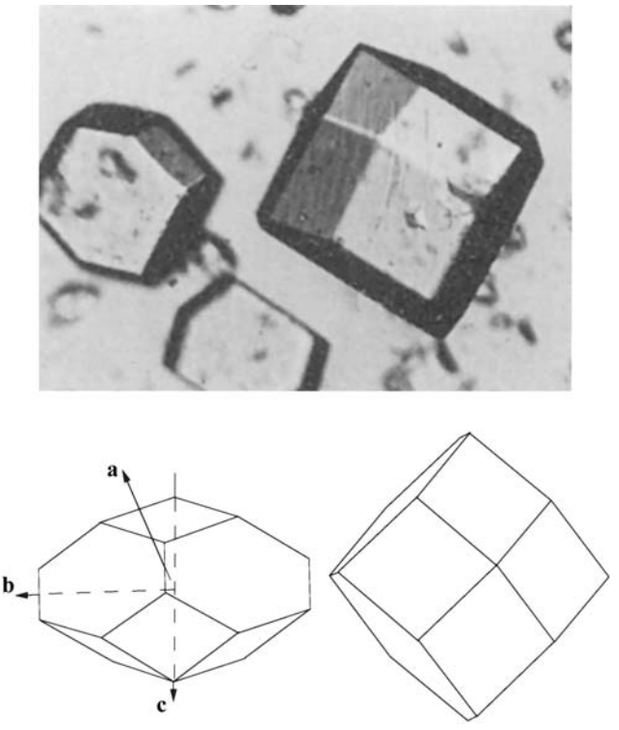

Fig. 26.1.2.1. Tetragonal lysozyme crystals with well developed {110} faces (left-hand crystal) and small {110} faces (right-hand crystal).

volumes in the range 0.5 to 1 mm^3, grew overnight or during the course of a few days.

The crystals were modified bipyramids with well developed {011} faces and bounded by hexagonal {110} faces that were developed to differing extents in individual crystals (Fig. 26.1.2.1). Many crystals grew in contact with the glass and were less regular in shape. Precession photographs confirmed that the unit-cell dimensions were $a = b = 79.1$ Å and $c = 37.9$ Å, and that the space group was $P4_12_12$ or $P4_32_12$ (Corey *et al.*, 1952). Each unit cell contained eight lysozyme molecules (one per asymmetric unit), molecular weight about 14 600, together with sodium chloride solution that made up about 33.5% of the weight of the crystal (Steinrauf, 1959).

With respect to structure analysis by the method of isomorphous replacement, these two enantiomorphic space groups have the great advantage of exhibiting three independent centrosymmetric projections, on the (001), (010) and (110) planes, corresponding to the *hk*0, *h*0*l* and *hhl* reflections, respectively. As a result, 173 of the 393 reflections from planes with spacings ≥ 6 Å have heavy-atom contributions exactly in or exactly out of phase with the protein contributions to the structure factors. This property greatly facilitated the determination and refinement of heavy-atom positions in the isomorphous derivatives used in the work on lysozyme.

26.1.2.3. *Preparation of heavy-atom derivatives*

For this work by RJP, lysozyme crystals were grown in glass vials in which a 1 ml solution contained between 20 and 50 mg of lysozyme. Attempts were then made to diffuse heavy-atom compounds into the pre-formed crystals. Stoichiometric amounts of heavy-atom compounds, such as K_2PtCl_4, $UO_2(NO_3)_2$, *p*-chloromercuribenzene sulfonate (PCMBS), *p*-chloromercuribenzoate (PCMB) and K_2HgI_4, were added to these vials, and precession photographs were taken of the crystals from each vial. The precession photographs of the putative derivative crystals were compared visually with those from the native protein by superimposing them on a light box. This showed immediately whether the cell dimensions had changed and if there were any significant changes in intensity. If these photographs showed no changes in intensity, then the amount of heavy-atom reagent was increased, and this process was continued until the crystals either showed intensity changes or disintegrated. When intensity changes were detected, the effect of increasing the concentration of heavy atom was explored with the object of establishing the optimum conditions for the preparation of derivatives with high occupancy of a small number of sites. In this way it was sometimes possible to follow decreases and increases of the intensities of weak reflections that went through zero with increasing concentration, and this indicated a reversal of the signs of the reflections from the native and derivative crystals.

In accordance with the example provided by the work on myoglobin, 9° precession photographs, which provide data to a resolution just beyond 6 Å, were used for these trials, and attention was concentrated on the [001] and [100] zones of reflections.

When these exploratory studies had produced a promising derivative, further precession photographs were taken for use in intensity measurements. For this purpose, two Ilford X-ray films were placed one behind the other in the camera cassette and they were exposed for about 24 h to Cu $K\alpha$ radiation from sealed X-ray tubes or for about 4 h to the same radiation from a rotating-anode tube. A second exposure with two films in the cassette was made for about 4 h with the sealed tube or 1 h with the rotating-anode tube in order to cover the full range of intensities that had to be measured. The intensity measurements were performed on the Joyce–Loebl recording densitometer, which scanned each row of reflections automatically but had to be moved manually from row to row. The heights above background of the peaks on the densitometer traces were measured with a ruler and these measurements provided the basic intensity data. Intensities recorded on the two films of each pair were brought to the same scale by calculation and application of a film transmission factor (usually between 2.5 and 3.0) and the corresponding factor relating films exposed for different lengths of time was obtained similarly. Weighted mean intensities were then calculated for each of the reflections. Lorentz–polarization (Lp) factor corrections were derived from a plot of the Lp factor against $(\sin^2\theta/\lambda^2)$, without consideration of asymmetric effects (Waser, 1951), and these factors were manually applied to the observed intensities to provide structure-factor measurements on an arbitrary scale.

The structure factors of the heavy-atom derivative crystals, $|F_{HP}|$, were scaled to those of the native protein, $|F_P|$, by a factor K, derived from the equation

$$K^2 \sum |F_{HP}|^2 = \sum |F_P|^2 S,$$

where S was a parameter, between 1.0 and 1.10, that depended on the assumed heavy-atom occupancy and was refined later in the process. Values of $\Delta F = K|F_{HP}| - |F_P|$ were then calculated for use in the determination of the heavy-atom positions.

26.1.2.4. *Determination of heavy-atom positions*

The high symmetry of the space group was greatly to our advantage, since the heavy-atom positions could be determined from difference-Patterson projection on the (100), (110) and (001) planes. In principle, all three coordinates of a heavy atom can be determined from projections on (100) or (110) alone. In practice, however, it was more straightforward to begin with the interpretation of the simpler projection on (001) before determining the z coordinate of the heavy atom from one of the other projections.

Apart from the effect of cross-over terms, which were sometimes detected as indicated above, these maps are effectively true

Patterson maps of the heavy-atom structures of the derivatives. These difference-Patterson maps were calculated on the MERCURY computer and even from time to time by the use of Beevers–Lipson strips – a less demanding task than it might appear, since only about 80 *hk*0 and 60 0*kl* reflections are included within the 6 Å limit.

26.1.2.4.1. *The mercuri-iodide (K_2HgI_4) derivative*

After trying several levels of substitution, RJP used the K_2HgI_4 salt at a molar concentration eight to ten times that of lysozyme. Based on the fact that lysozyme contains two methionine residues per molecule, and in keeping with a suggestion of Bluhm *et al.* (1958), RJP was expecting to see two heavy-atom sites,* but the *hk*0 difference-Patterson map was interpretable in terms of a single site of substitution. This site, however, was very close to the crystallographic twofold axis that runs along a diagonal in the [001] projection of the unit cell, and it proved necessary to correct the details of the first interpretation when phase information became available from other derivatives. It then appeared that there was one HgI_4^{2-} (or HgI_3^-) on the twofold axis between two protein molecules, but that it was best modelled by two closely spaced sites to allow for the elongated shape of the group (see Table 26.1.2.1). Several other heavy-atom salts, including K_2HgBr_4, K_2PtBr_4 and K_2AuCl_4, gave derivatives in which the heavy atom was attached to the same site as K_2HgI_4, and consequently seemed not to provide useful additional phase information.

26.1.2.4.2. *The palladium chloride (K_2PdCl_4) derivative*

An attempt to use K_2PtCl_4 to produce a useful derivative gave disordered crystals, but a substitute for it was found by soaking crystals in K_2PdCl_4 at a molar ratio of 3:1 relative to lysozyme. Despite the relatively light Pd atom, this derivative gave an easily interpretable difference-Patterson map (see Fig. 26.1.2.2) that yielded very good *R* factors,

$$R = \sum \| F_{HP} | - | F_P \| - | F_H(\text{calc}) | / \sum | F_H(\text{calc}) |,$$

where the summations are over centric reflections only.

26.1.2.4.3. *The o-mercurihydroxytoluene p-sulfonate (MHTS) derivative*

The *hk*0 difference-Patterson map of the *p*-mercuribenzene sulfonate (PCMBS) derivative was interpretable in terms of a single site of substitution at 8 Å resolution, but it was not useful beyond about 8 Å because of lack of isomorphism. RJP and RHF then explored the usefulness of MHTS as a derivative. This compound had been specially synthesized by JWHO in the hope that a small rearrangement of groups present in PCMBS would lead to an isomorphous derivative. Happily, this strategy worked, and MHTS gave a useful isomorphous derivative in which the major site overlapped that of PCMBS (Fig. 26.1.2.3).

26.1.2.4.4. *Other potential derivatives*

As is usual in protein work, RJP tried many other heavy-atom compounds (Poljak, 1963), but none gave useful results. In particular, a uranyl derivative, obtained by the use of UO_2NO_3, gave difference-Patterson maps that were difficult to interpret, and it was not taken further at this stage.

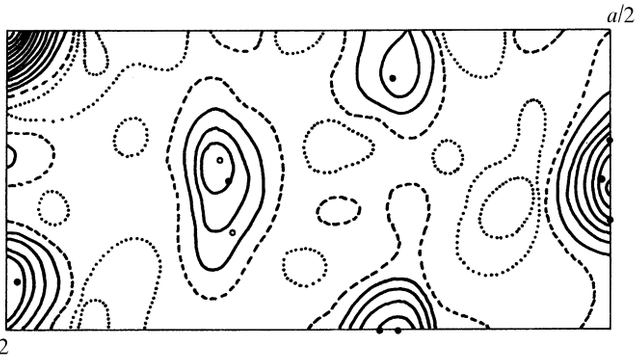

Fig. 26.1.2.2. Difference-Patterson *h*0*l* projection map for the derivative obtained with K_2PdCl_4. The ends of heavy-atom double-weight (filled circles) and single-weight (open circles) vectors are shown (R. J. Poljak, unpublished material).

26.1.2.5. *Refinement of heavy-atom parameters*

Refinement of the heavy-atom parameters was first performed by the use of Rollett's (1961) least-squares program on the MERCURY computer, using the $|\Delta F|$ values as structure amplitudes. This procedure gave satisfactory results for the K_2HgI_4, K_2PdCl_4 and MHTS derivatives described above, and they were used, therefore, in an attempt to determine the structure of the protein to 6 Å resolution in three dimensions.

26.1.2.6. *Analysis in three dimensions*

26.1.2.6.1. *X-ray intensity measurements*

We had three options for the collection of three-dimensional data. First, we could have used precession photographs and

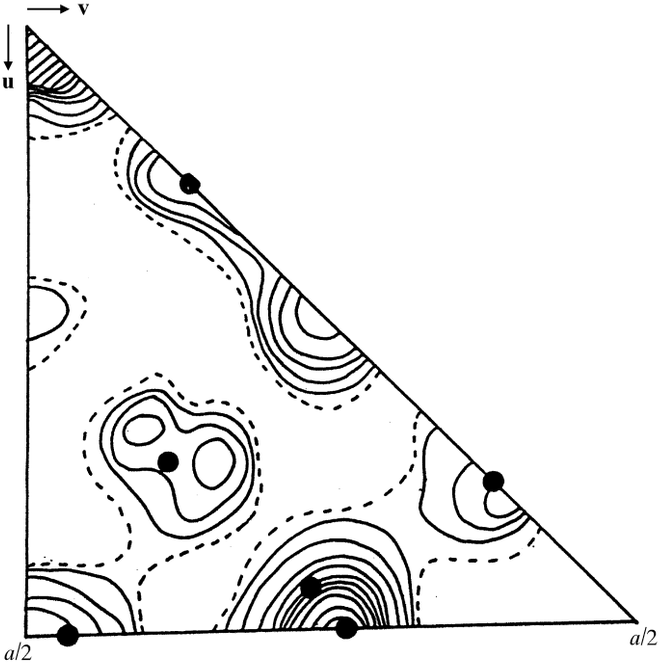

Fig. 26.1.2.3. Difference-Patterson *hk*0 projection for the derivative obtained with MHTS. The large peak at $\frac{1}{4}, \frac{1}{4}$ is not explained by this solution (Fenn, 1964).

* In order to study the nature of the ligand formed from K_2HgI_4, RHF and DCP studied the structures of two compounds, $(CH_3)_3S{\cdot}HgI_3$ and $[(CH_3)_3S]_2{\cdot}HgI_4$ (Fenn *et al.*, 1963; Fenn, 1964), which were prepared for this purpose in the laboratory by JWHO. These studies showed that HgI_3^- and HgI_4 are, respectively, planar trigonal and tetrahedral in configuration. Meanwhile, Dr Helen Scouloudi was examining the nature of the K_2HgI_4 derivative of seal myoglobin (Scouloudi, 1965) and showed that the ligand was HgI_3^- and not associated with methionine.

densitometry, as in the study of myoglobin (Kendrew *et al.*, 1960). Second, the manually controlled three-circle diffractometer used to make some of the measurements in the 6 Å study of myoglobin (Kendrew *et al.*, 1958) and haemoglobin (Cullis *et al.*, 1962) was available and, third, there was the prototype linear diffractometer (Arndt & Phillips, 1961). We chose the last option, since CCFB, RHF and DCP were well experienced in using the instrument, and it offered the opportunity of measuring the protein reflections automatically and in a relatively short time compared with the other methods. Before he went on leave to MIT, ACTN had written a computer program for the University of London MERCURY computer to process the diffractometer data (North, 1964) and, on his return in September 1961, he readily accepted an invitation to join the team to continue with this and other related aspects of the work.

The design of the linear diffractometer was based directly on the reciprocal-lattice representation of the genesis of X-ray reflections. The principle is illustrated in Figs. 26.1.2.4(*a*) and (*b*), which show the familiar Ewald construction. YXO represents the direction of the incident X-ray beam with X the centre of the Ewald sphere and O the origin of the reciprocal lattice. A′OA, B′OB and C′OC are the principal axes of the reciprocal lattice, here assumed to be orthogonal. XP is the direction of the reflected X-ray beam corresponding to the reciprocal-lattice point P, which lies on the surface of the sphere of reflection. The reciprocal lattice can be rotated about the axis C′OC, and this axis can be inclined to the direction of the incident X-ray beam by rotation about the axis D′OD, which is perpendicular to the incident beam.

The linear diffractometer was simply a mechanical version of this diagram. The reciprocal lattice was represented by three slides, A, B and C, which were parallel, respectively, to A′OA, B′OB and C′OC. They were mounted to rotate about the axis C′O and arranged so that the saddle P could be set to any position in space within the coordinate system that they defined. This saddle P was connected to the point X by means of a link of fixed length, XP = XO, corresponding to the radius of the sphere of reflection. The link XP always lay along the direction of the reflected X-ray beam and thus became the counter arm of the diffractometer. The crystal was mounted at X for rotation about the axis R′XR (independent of the link XP, which pivoted about an independent coaxial bearing at X). The rotation of the crystal about this axis was coupled by means of gears, pulleys and steel tapes to the rotation of the slide system about the axis C′OC. The axes R′XR and C′OC, held parallel by means of parallel linkages, could be tilted with respect to the incident X-ray beam by rotation about the axes D′OD, E′XE, as shown in Fig. 26.1.2.4(*b*).

The scale of the instrument clearly depended only on the length chosen for XP = XO. In the instruments used in the lysozyme work, this length, which is equivalent to one reciprocal-lattice unit, was five inches. The position of the saddle P on the three slides was controlled by means of lead screws, all of which were cut with 20 turns per inch. Hence the counters, which indicated revolutions and fractions of a revolution of the lead screws, read directly in decimal divisions of reciprocal-lattice units. The screws in slides A and B were driven by means of synchro-receiver motors, forming a synchro link with corresponding transmitters in the control panel. Slide C was set manually, together with the inclination angle μ, for the measurement of upper-level reflections in the Weissenberg equi-inclination mode, Fig. 26.1.2.4(*b*).

The coupling of the rotations of the crystal and reciprocal lattice about the axes R′XR and C′OC, respectively, was interrupted by two ancillary mechanisms. The first simply allowed for independent rotation of the crystal with respect to the slide system and was used for setting the reciprocal-lattice axes in the equatorial plane parallel to slides A and B, and for any fine adjustment of the crystal rotation that might be necessary during the measurement procedure. The

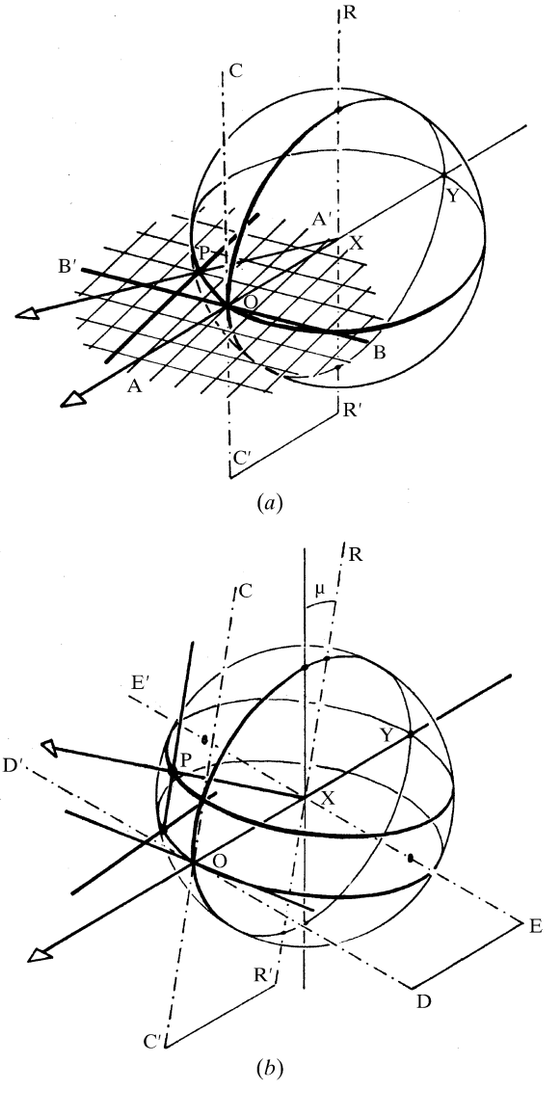

Fig. 26.1.2.4. Reciprocal-space diagrams showing the direction of the incident X-ray beam, the Ewald sphere and the genesis of a reflection (*a*) in an equatorial plane and (*b*) in the equi-inclination setting. Principal reciprocal-lattice directions are shown as thick lines. They also represent the slides in the diffractometer. The rotation of the diffractometer slide system about the axis C′OC is coupled to the rotation of the crystal about the axis R′XR by gears, pulleys and steel tapes. The counter arm of the diffractometer is represented by the fixed link XP = XO. Reproduced with permission from Arndt & Phillips (1961). Copyright (1961) International Union of Crystallography.

second interruption consisted of a mechanism for oscillating the crystal about the position for any reflection while X-ray intensity measurements were made. This oscillation mechanism (Arndt & Phillips, 1961) rotated with the crystal as the diffractometer was being set to a reflection position, and then controlled the independent motion of the crystal for the measurement of the integrated intensity of the reflection. The crystal remained stationary at a given angular setting for time t, was rotated at a uniform rate over a predetermined angular range for a time $2t$, remained stationary at the final angular setting for a further time t, and then returned quickly to its original setting. The correct setting for the reflection peak was at the midpoint of the rotation, which might be set to be through any angle from 1 to 5°. For initial adjustments, the motor could be arrested at this midpoint by means of a micro-switch operated by a switching disc rotating with the

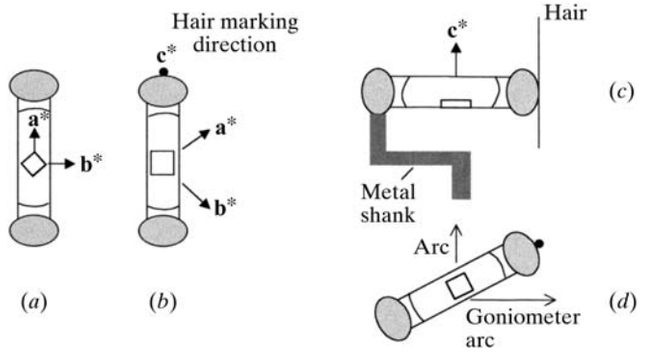

Fig. 26.1.2.5. Crystal mounting. (*a*) Rotation about the a^* axis; (*b*) rotation about [110], preliminary to c^* mounting; (*c*) rotation about the c^* axis (elevation); and (*d*) rotation about the c^* axis (plan).

cam. This disc otherwise actuated contacts that started and stopped the intensity measurements.

The X-ray intensities were measured with a side-window xenon-filled proportional counter made by 20th Century Electronics, together with associated amplifiers and pulse-counting circuits (Arndt & Riley, 1952). The proportional counter had a high quantum efficiency for the measurement of Cu $K\alpha$ radiation (about 80%) and, when the operating potential and pulse-height discriminating circuits were carefully set, it provided useful discrimination against radiation of other wavelengths. The output from the proportional counter and its associated circuitry was fed directly to a teleprinter, which gave both a plain-language print out and a five-hole punched paper tape for input to the computer. Each count was provided with a check digit derived by a 'ring-of-three' circuit, wired in parallel with the main electronic counter. During data processing, the check digit was compared with the count *modulo* 3: inequality of the two numbers was taken to indicate an error in the counting circuit.

Three counts were made: the first was a background count n_1, made while the crystal was stationary on one side of the reflection position; the second was an integrated intensity count N, accumulated as the crystal rotated through the reflection; and the third was a further background count n_2. The background-corrected integrated intensity of the reflection was taken to be

$$N_o = N - (n_1 + n_2).$$

At this stage of the work, measurements were made of one reciprocal-lattice level at a time in the equi-inclination mode that has no blind region at the centre. In each level, the diffractometer was driven to a reflection *hkl* (for example) at the limit of resolution of the data set to be collected. The diffractometer then moved in a series of equal steps along the scanning slide. At the end of each step, the oscillation mechanism took control for the measurement of intensity and background. After each measurement, a further step was taken on the scanning slide, and the process continued until a limit switch, set to the required resolution limit, was reached. The diffractometer then completed the current translation, measured the last reflection in that row, and then moved one step on the stepping slide to the next parallel row. This row was scanned, in the opposite direction, until the limit switch was reached again. In this way, the whole of a reciprocal-lattice level could be measured. In order to change to another level, the vertical slide C and the inclination angle μ had to be manually adjusted.

This account ignores two difficulties, one inherent in the design of the diffractometer, and the other specific to the lysozyme crystals. First, the instrument required a good deal of supervision, since it did not set itself very well for the measurement of low-angle reflections. Second, the crystals were not easily mounted so that the *c* axis, the most convenient axis for efficient data collection, since it is perpendicular to the most densely populated reciprocal-lattice planes, coincided with the crystal-rotation axis of the diffractometer.

The first problem was overcome by efficient teamwork and was much eased by the fact that RHF assumed responsibility for the MHTS derivative as part of her PhD work; the second was solved by making most of the measurements from crystals mounted to rotate about the [100] axis. These crystals were oriented so that the b^* and c^* axes were parallel to the horizontal slides of the diffractometer, and the measurements were made in levels of constant H by scanning along rows parallel to b^* and stepping to adjacent rows along c^*. A number of reflections could not be measured in this way, however, because reflections near the a^* axis were too broad to measure, particularly in the upper reciprocal-lattice levels. This difficulty was overcome by mounting some crystals with the *c* axis of the tetragonal crystals perpendicular to the length of the capillary tube, with the [110] axis parallel to the tube. These specimens were then mounted on a right-angled yoke so that the capillary tube was perpendicular rather than parallel to the goniometer axis (Fig. 26.1.2.5).

Given the morphology of the crystals, with an essentially square habit bounded by {110} faces (Figs. 26.1.2.1 and 26.1.2.5), all the reflections in the *hkl* octant could be measured in levels with constant L values without inclining the capillaries by more than about 40° to the X-ray beam. The horizontal slides of the diffractometer were set to be parallel to the a^* and b^* axes of the crystal. Some quadrants of *hkL* reciprocal-lattice levels were then scanned along rows parallel to the a^* axis and stepped along the b^* axis. Enough measurements were made in this mode to cover the 'blind' region in the *Hkl* levels and provide an appropriate number of intersecting levels for scaling all the measurements into a consistent set.

Care was taken during all these measurements to index the reflections in a right-handed system of axes. Given the transparent relationship between the slide system of the diffractometer and the crystal geometry, this was easily accomplished, and it was necessary for the subsequent use of anomalous scattering (Bijvoet, 1954) in the phase determination.

During the measurements from the native and derivative crystals mounted for rotation about [100], the variation in peak intensity of the 200 reflection with φ, the angle of rotation about the axis C′OC (Fig. 26.1.2.4*a*), was also recorded (Fig. 26.1.2.6). (200 is the lowest-order reflection available for this purpose in this space group.) These records were then used in the data-processing stage to correct the measurements for absorption by the method described by Furnas (1957). According to this method, the absorption suffered

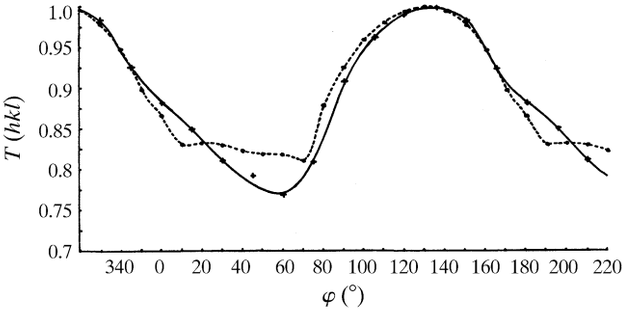

Fig. 26.1.2.6. Absorption curve. Variation of relative transmission $T(hkl) = I(\varphi_{hkl})/I_{max}(\varphi)$ $[= 1/A(hkl)]$ with rotation angle φ for the 200 reflection and the crystal rotating about the normal to (200). Solid line: measured curve; broken line: calculated curve, neglecting effect of mother liquor and capillary. Reproduced with permission from North *et al.* (1968). Copyright (1968) International Union of Crystallography.

by the incident and reflected beams for any reflection *hkl* is indicated by the relative intensity of 200 when the X-ray beams are parallel to the reflecting planes *hkl*. This is a fair approximation for reflections at low angles. This method could not be used, however, for the measurements made from crystals mounted to rotate about the [001] axes, since a full rotation about this axis was not possible. These measurements were not corrected for absorption errors.

Finally, the diffractometer was reset manually at regular intervals during data collection to measure the intensities of a number of reference reflections. These measurements were used to monitor the stability of the system and the extent of irradiation damage to the crystal, and they were also recorded on the paper-tape output for analysis by the data-processing program. An attempt was made to minimize irradiation damage by using a shutter to expose the crystal only during the measuring cycle.

26.1.2.6.2. *Data processing*

The measurements N, n_1 and n_2, together with the indices of the reflections, *hkl*, were all printed out in plain language on a teleprinter and punched in paper tape for direct transfer to a computer (Fig. 26.1.2.7). The plain-language record was important during measurement of the low-angle reflections, when the diffractometer had to be adjusted by hand. Not all imperfections in the measurements were easily spotted at this stage, however, and ACTN's data-processing program (North, 1964) therefore incorporated systematic checks on the quality of the measurements.

The program checked for the following contingencies:

(1) malfunction of the diffractometer-output mechanism leading to the paper tape being an inaccurate record of the measurements, generally because the tape punch had failed to perforate the tape or had 'stuttered';

(2) errors by the pulse counters, detected by the 'ring-of-three' circuit;

(3) peak counting rate so high that counting-loss errors were appreciable;

(4) count on reflection not significantly above background;

(5) failure of diffractometer to set crystal or counter correctly; and

(6) gradual drift in the experimental parameters, including movement of the crystal within its mounting and irradiation damage to the crystal.

These checks were made while the diffractometer tape was being read into the computer, and a monitor output was produced simultaneously, as shown in Fig. 26.1.2.8. The checks depended in large part on the fact that the significance of an intensity measurement may be assessed in terms of counting statistics. The standard deviation of a background-corrected count, N_o $(= N - n_1 - n_2)$, is given by $\sigma^2(N_o) = N + n_1 + n_2$, and the ratio $\sigma(N_o)/N_o$ may be taken as an indication of the significance of the measurement. Measurements were rejected when this ratio exceeded unity. N_o might then have been taken as zero but, following Hamilton (1955), we considered it preferable to replace N_o by a fraction (0.33 for centric and 0.5 for acentric reflections) of the minimum background-corrected count that we should have considered acceptable. Reflections were treated in the same way whether the net count N_o was positive or negative, but measurements were rejected if N_o was negative and $|N_o| < \sigma(N_o)$.

Mis-setting of the crystal was frequently revealed by marked inequality of the background counts. Measurements were therefore rejected if the difference between the two backgrounds exceeded three or four standard deviations, that is if $(n_1 - n_2)^2 > b^2(n_1 + n_2)$, where b is the appropriate constant.

After monitoring the quality of the data in this way, the program proceeded: (i) to extract background-corrected counts; (ii) to apply a correction for irradiation damage derived from any systematic variation in the intensities of the reference reflections; (iii) to sort the reflections into a specified sequence of indices; (iv) to apply Lorentz–polarization factors; and (v) to apply absorption corrections (the data for which were read separately from a specially prepared punched tape, Fig. 26.1.2.6). The outputs from this program comprised data sets from a number of individual crystals of the native protein and the three derivatives. The scale factors needed to bring the measurements from the individual crystals of

```
*02+13+00+
:000234 002500 000284

*01+13+00+
:000260 001660 000279

*00−13+00+
:000181 000522 000212

*00−12+00+
:000225 000651 000236

*01+12+00+
:000327 005860 000362

*02+12+00+
:000281 006717 000260

*03+12+00+
:000236 000507 000240

*04+12+00+
:000225 000488 000241

*05+12+00+
:000244 003763 000263
```

(*a*)

```
*    7    2
   0   4912   98
   1   4836   97
   2    276   33
   3  12710  127
   4   1216   49
   5    802   40
   6    398   32
   7   1551   47
   9    935   37
  10   1257   38

     8    2
   0   9855   99
   1   1450   44
   2   6117   61
   3   4733   95
   4     61   26
   5    395   32
   6   1304   39
   7    129   28
   8    183   27
   9   1725   52
```

(*b*)

Fig. 26.1.2.7. Typical output from the linear diffractometer. (*a*) Indices h, k, l followed by background (n_1), peak (N), background (n_2) counts. (*b*) Listing ready for the next stage in data processing with indices * h k followed by l, background corrected peak and standard deviation. Reproduced with permission from North (1964). Copyright (1964) Institute of Physics.

```
PBDN 3/2
TLY  PDBR 4  D 2 * 1/5    K = 1 2 : 1 4 : 1 6    24/2/64
 01   01   14 :=  SD  966
 01   00   14 PE 2
 01  −01   14 :−  SD  999  :=  SD  156
 01  −04   14 :−  BG  :+  BG
 01  −13   14 :=  N−  SD  578  BG
 01  −13   14 :−  N−  :=  N−  BG
 01  −13   14 :−  SD  391  :=  N−  SD  234
 01  −14   14 :−  SD  132  :+  SD  216
 01  −15   14 :=  SD  813
 00  −15   14 :−  BG  :=  BG  :+  BG
 00  −12   14 :=  BG
 00  −10   14 :+  BG
 00  −10   14 :=  BG
 00  −04   14 :=  BG
 00  −04   14 :+  BG
 00   04   14 :−  BG  :=  BG  :+  BG/

TOTAL  PE1  PE2  PE*  SD  N−  BG  N>H
  42    0    1    0   9   4  15    0
```

Fig. 26.1.2.8. Format of monitor output in which the computer lists reflections that fail the tests for format or significance. PE1 signifies punching error, indices; PE2, punching error, measurements; PE3, failure of electronic check on counting circuits; SD, standard deviation greater than set limit; N-, net count negative; BG, backgrounds significantly different; N > H, gross counts exceed counting-loss limit. This output was from the version of the program designed to be used with the diffractometer fitted with three counters. The symbols: −, =, + refer, respectively, to reflections measured by the lower, central and upper counters. Reproduced with permission from Arndt *et al.* (1964). Copyright (1964) Institute of Physics.

each crystal species to a common scale were determined from the intensities in common rows by the use of a program written by Rollett (Rollett & Sparks, 1960). The scaling factors were applied and the data merged, a weighted-average intensity being determined when more than one estimation was available.

26.1.2.6.3. *The absolute scale of the intensities*

An attempt was made to determine the absolute scale of the measured intensities by comparison with the intensities diffracted by anthracene, a small organic crystal of known structure. This method had worked well in a determination of the absolute scale for seal myoglobin (Scouloudi, 1960), but it did not give a satisfactory result with lysozyme, mainly because of the difficulty of measuring the crystal volumes precisely enough. Accordingly, we used Wilson's (1942) method to provide an estimate of the absolute scale of the intensities, knowing very well that it does not give an accurate estimate for protein data, especially at low resolution. Nevertheless, this scale gave reasonable values for the occupancies of the heavy-atom sites.

26.1.2.6.4. *Re-assessment of heavy-atom derivatives*

Given the three-dimensional data to 6 Å resolution for the native crystals and the three derivatives, it was next possible to calculate three-dimensional difference Patterson maps for the derivatives using the terms

$$|\Delta F|^2 = ||F_{PH}| - |F_P||^2$$

as coefficients in the Fourier series. This synthesis, which is now well known in protein-structure analysis, gives a modified Patterson

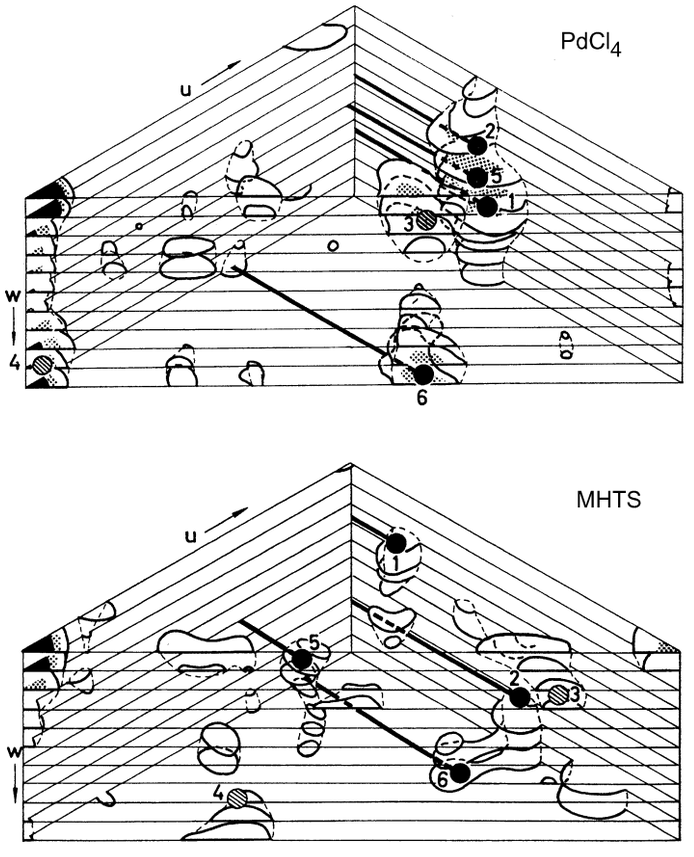

Fig. 26.1.2.9. Three-dimensional $|\Delta F|^2$ syntheses for the $PdCl_4$ and MHTS derivatives. The positions of peaks in the Harker sections are marked and numbered (Fenn, 1964).

of the heavy-atom structure in which the heavy-atom vectors appear at reduced weight in a complex background (Blow, 1958; Phillips, 1966). Nevertheless, the $|\Delta F|^2$ maps for $PdCl_4$ and MHTS were readily interpreted in terms of single heavy-atom substitutions, particularly because the vectors involved were all confined to defined Harker sections (Fig. 26.1.2.9).

The map of the HgI_3^- derivative was not so satisfactory, and this led to the discovery, during refinement of the heavy-atom parameters, that the mercury occupancy declined as a function of irradiation time. Consequently, the HgI_3^- data from two individual crystals were treated separately during the remaining stages of the analysis.

The least-squares refinement mentioned in Section 26.1.2.4 was not wholly satisfactory in that it included no provision for refining heavy-atom occupancy. Accordingly, DCP – with some help from ACTN – wrote a computer program for the MERCURY computer based on Hart's (1961) method, in which the heavy-atom positions, occupancies (O) and temperature factors (B) were refined simultaneously together with the scale factors (S) between heavy-atom and native structure factors. All the centric reflections from the [100], [110] and [001] zones were included in this refinement.

The quantity minimized for each derivative was

$$R' = \sum(||F_{PH}| - |F_P|| - |F_H|)^2,$$

where F_H is the calculated heavy-atom contribution to the derivative structure factor. The method involves calculating R' for each parameter at its current value p_n and at values $p_n \pm \Delta p_n$ and $p_n \pm 2\Delta p_n$, where each of the parameters is shifted in turn, the shifts having been specified, while the other parameters are at the unshifted value. Thus for each of the parameters (x, y, z, O, B and S), four values of R' are obtained for the shifted values plus the value of R' for the unshifted parameters, the latter being denoted ε_u.

Let the minimum value of R' from all the calculations be $\varepsilon_{\min}$ and for the parameter p_n, the minimum of the list of five values of R', be ε_{qn}, corresponding to the value q_n of p_n. Then, according to the method of steepest descents, the shift to be applied to the parameter p_n is

$$(q_n - p_n)(\varepsilon_u - \varepsilon_{qn})/(\varepsilon_u - \varepsilon_{\min}).$$

If $\varepsilon_{qn} = \varepsilon_u$, that is, if the unshifted value of the parameter gave the minimum value of R', then the shift was divided by 4 for the next cycle. Otherwise the shift was kept constant. Thus the new parameters and shifts were determined for the next cycle of refinement, and the process was repeated until convergence.

This program worked well, and RJP, who was reading *Candide* at the time, named it *Pangloss* – it gave the best possible values for the heavy-atom parameters. These values, which include two separate sets for the HgI_3^- derivative, are shown in Table 26.1.2.1. At this stage, an important check was carried out. The coordinates of the heavy-atom site in each derivative were referred to an origin at the junction of a twofold axis and a twofold screw axis. However, there are four such intersections in the unit cell and, in order to ensure that the same origin had been chosen for each derivative, the sign predictions for the centric reflections from each derivative – which were checked by hand throughout this exploratory stage – were compared. They agreed well, thus establishing that the choice of origin was the same for each derivative.

26.1.2.7. *Phase determination at 6 Å resolution*

These parameters were used to determine the phases of the protein reflections. A proportion of these phases were first determined by the graphical method suggested by Harker (1956), which had been used in the 6 Å stage of sperm-whale myoglobin (Kendrew *et al.*, 1958). We treated the process as a group activity in which different individuals took responsibility for reading out the

Table 26.1.2.1. *Heavy-atom parameters used in the final phase calculation for the lysozyme structure*

E is the average difference between observed and calculated heavy-atom changes of centric reflections (electrons); *R* is the reliability index for observed and calculated heavy-atom changes of centric reflections and *R*′ is the Kraut (Kraut *et al.*, 1962) agreement index for all reflections.

			Mercuri-iodide			
	$PdCl_4$	MHTS	Crystal 1*		Crystal 2†	
x	0.147	0.218	0.131	0.178	0.125	0.178
y	0.841	0.620	0.869	0.822	0.875	0.822
z	0.963	0.054	0.250	0.250	0.250	0.250
Occupancy (e)	72	47	74	74	72	72
B (Å^2)	52	−18	38	38	111	111
E (e)	57	66	66		55	
R (%)	35	49	42		39	
R′ (%)	10.7	11.6	11.6		12.2	

* Centric 0, 1, 2*kl* reflections only.
† *hk*0 reflections only.

various Harker components [F_P, F_{PH}, F_H(calc) *etc.*] for each reflection in turn. This was a useful bonding exercise and improved our familiarity with the rather cosmopolitan accents in use within the group. Scientifically, it was also an encouraging experience since it showed that reasonably consistent results could be obtained from the three different derivatives, *and* that the anomalous-scattering measurements were capable of making a significant contribution to the phase determination. It soon became clear that the most significant anomalous contributions had to be included in such a way as to retard the phase of the derivative structure factor if these indications were to agree with the phase predictions derived from the isomorphous differences alone (Fig. 26.1.2.10).

The implication of this observation was that the space group is $P4_32_12$ rather than $P4_12_12$, that is, the fourfold screw axis is left handed.

These results encouraged us to go ahead with the computer calculation of phases by the phase probability method applied to the isomorphous differences (Blow & Crick, 1959; Dickerson *et al.*, 1961) and to the anomalous-scattering differences (Blow & Rossmann, 1961). This program was written by ACTN and it was used first to confirm the space-group identification. Overall, the mean figure of merit obtained with the anomalous contribution consistent with $P4_32_12$ was somewhat higher than the alternative, though to a lesser extent than we had anticipated. This observation led at a later stage to reconsideration of the way in which the anomalous scattering was incorporated in the phase determination (see Section 26.1.3.9).

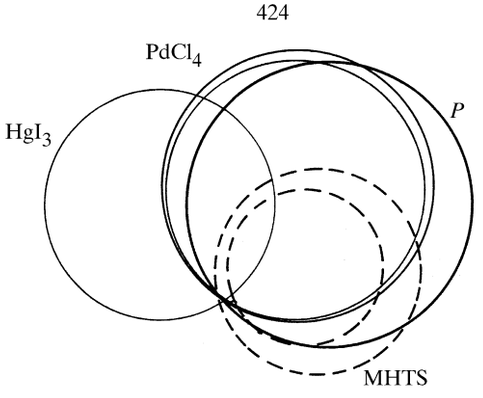

Fig. 26.1.2.10. Phase determination for reflection 424. In this Harker diagram, the heavily traced protein circle (with radius $|F_P|$) is labelled *P*. The circles with radii $|F_{HP}|$ obtained with $PdCl_4$, MHTS and HgI_3 are shown. Values of the anomalous-scattering pairs *hkl* and *khl* were used for $PdCl_4$ and MHTS. The position of the protein vector, weighted by its figure of merit, is shown as a small open circle (R. J. Poljak, unpublished results).

The quality of the phase determination was indicated by the figure of merit. For the acentric reflections, this was 0.79, an encouraging result since it compared favourably with that obtained in the low-resolution study of haemoglobin (Perutz *et al.*, 1960). The mean figure of merit for the centric reflections, on the other hand, was 0.95, so that the overall value was 0.86. A check on the sign predictions for centric reflections showed that the three derivatives gave satisfyingly similar results.

These phases were also used to calculate difference-Fourier maps, showing the heavy atoms in the derivatives by the use of coefficients

$$|\Delta F| = ||F_{PH}| - |F_P||$$

associated with the protein phases and weighted by the figures of merit of each phase determination. These difference maps in three dimensions are shown in Fig. 26.1.2.11.

They revealed the presence of small subsidiary sites in these two derivatives, but these minor sites were not taken into account.

26.1.2.8. *The electron-density map of lysozyme at 6 Å resolution*

The electron-density distribution was calculated on the MERCURY computer, by means of a program written by Owen Mills that was in general use at the time, with structure amplitudes weighted by the figures of merit so as to give the 'best' Fourier (Blow & Crick, 1959). The map was contoured by hand and plotted on clear plastic (Perspex; Plexiglass in the USA) sheets (Fig. 26.1.2.12).

The first objective in studying this map was to determine the boundary of a single molecule. Comparison of the unit cells of various crystal forms of lysozyme (Steinrauf, 1959) suggested that in tetragonal lysozyme, the molecule occupies the full length of the *c* axis and, on average, one-eighth of the *ab* plane. On this basis, the map of Fig. 26.1.2.12 includes the whole of the *c* axis and a sufficient area of each section to ensure that one whole molecule is included in addition to parts of neighbouring molecules. Only contours indicating where the electron density is greater than average are included.

Some featureless regions of average electron density (0.4 e Å^{-3}) were immediately apparent. The most marked of these was around the twofold screw axis parallel to **c**. This axis is intersected at intervals of 9.5 Å by twofold rotation axes, and packing considerations, therefore, made it impossible for substantial parts of the molecule to penetrate into the neighbourhood. Similarly, the immediate vicinity of the fourfold screw axis was also without significant features and was clearly a region of intermolecular space. The twofold rotation axes also helped to determine the boundary, particularly where relatively high density approached or intersected them, as it did in two places. It was clear that such regions must represent close contacts or bridges between adjacent molecules.

Following the example of the haemoglobin study (Perutz *et al.*, 1960), we decided at this stage to make a balsa-wood model of the electron density to help visualize the molecule. Instead of producing a stack of sections through the molecule, however, CCFB devised a way of shaping the sections to make a smooth model of the volume

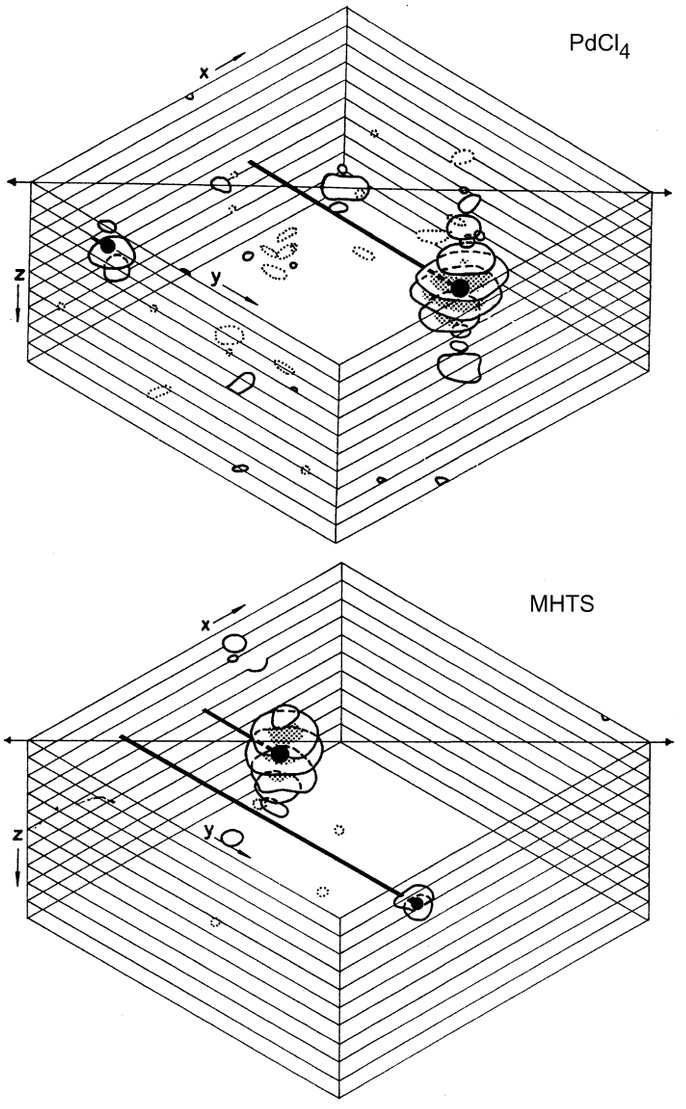

Fig. 26.1.2.11. Electron-density difference syntheses showing the main sites of substitution and subsidiary sites with low occupancy in the $PdCl_4$ and MHTS derivatives (Fenn, 1964).

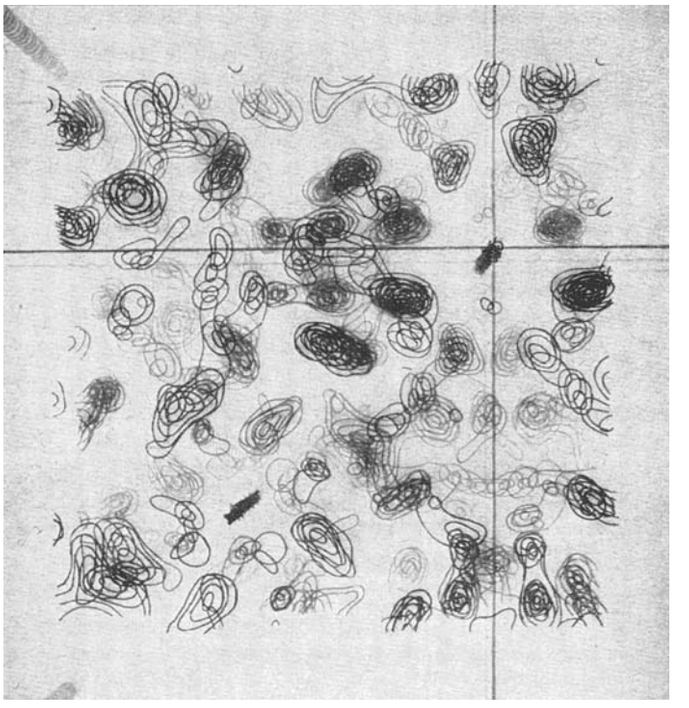

Fig. 26.1.2.12. Electron-density distribution in lysozyme at 6 Å resolution viewed parallel to the *c* axis. The horizontal and vertical lines represent the twofold rotation axes and intersect the twofold screw axis, upper right of centre. The fourfold screw axis is at the lower left of centre. The contour interval is 0.07 e Å^{-3}, the lowest heavy contour being at 0.6 e Å^{-3}. The absolute scale is approximate. Reproduced with permission from *Nature* (Blake *et al.*, 1962). Copyright (1962) Macmillan Magazines Limited.

occupied by electron density greater than about 0.53 e Å^{-3}. The result is shown in Fig. 26.1.2.13.

One asymmetric unit is shown in white, with additional pieces in grey to show alternative shapes. The white pieces together make up the most compact asymmetric unit, which is roughly ellipsoidal in shape with axes 52 × 32 × 26 Å. Later work showed that this asymmetric unit represented a single molecule but, at this stage, we were scrupulous in detailing the alternative interpretations. There is a region of low density that divides the model roughly into two halves, and although we speculated about this we refrained from making any suggestions about its possible significance in our description of the structure (Blake *et al.*, 1962). Instead, we noted that the two halves of the model could be assembled differently, following the crystal symmetry, so as to form a dumb-bell shaped molecule connected at PP′ (Fig. 26.1.2.13*d*).

Our second objective was to determine as far as possible the course of the polypeptide chain and the positions of the disulfide bridges. This proved to be impossible. In comparison with the maps of myoglobin (Kendrew *et al.*, 1958) and haemoglobin (Perutz *et al.*, 1960) at this resolution, it was immediately apparent that this map of lysozyme had a much smaller proportion of clear-cut rod-like features representing α-helices. This was not a surprise, since optical rotatory dispersion measurements (Yang & Doty, 1957) suggested that only 30–40% of the polypeptide chain in lysozyme is in the form of α-helix, as compared with 77% in myoglobin. In addition, Hamaguchi & Imahori (1964) had distinguished the presence of a region of β-sheet in lysozyme before completion of the X-ray analysis. The task of tracing the polypeptide chain, which was difficult with myoglobin, was impossible with lysozyme, since the connectivity of the non-helical regions was often not discernible. The existence of four disulfide bridges, which were expected to have about the same electron density as helices at this resolution, complicated the problem further.

Accordingly, we concluded that defining the shape of the molecule and its tertiary structure would have to await further studies at higher resolution. Meanwhile, Corey and his colleagues (Stanford *et al.*, 1962) and Dickerson *et al.* (1962) published interim accounts of their work at the same time as our work was published (Blake *et al.*, 1962).

At this stage, in the autumn of 1962, RJP left for three months for the MRC Laboratory in Cambridge and then joined Howard Dintzis at Johns Hopkins. At about the same time, DFK joined the team to continue the analysis to high resolution, and LNJ joined DCP as a graduate student and began work related to the activity of the enzyme.

26.1.3. Analysis of the structure at 2 Å resolution

The structure of chymotrypsinogen (Kraut *et al.*, 1962) at 6 Å resolution was published a few months before the corresponding work on lysozyme. Compared with the work on the globins,

references