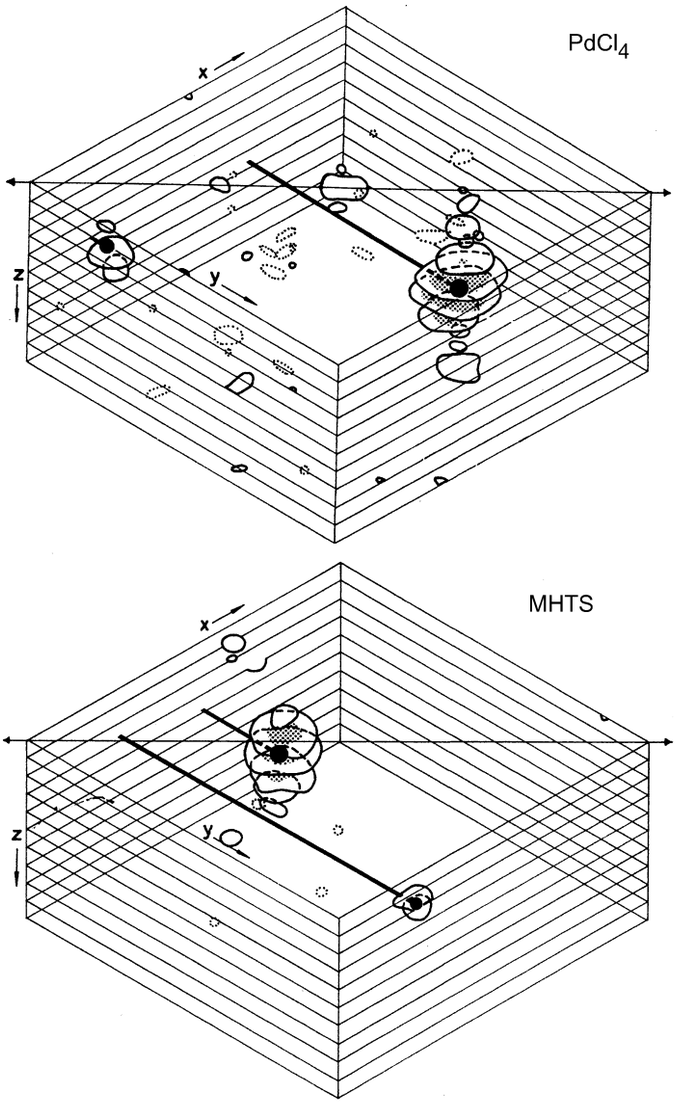

Fig. 26.1.2.11. Electron-density difference syntheses showing the main sites of substitution and subsidiary sites with low occupancy in the $PdCl_4$ and MHTS derivatives (Fenn, 1964).

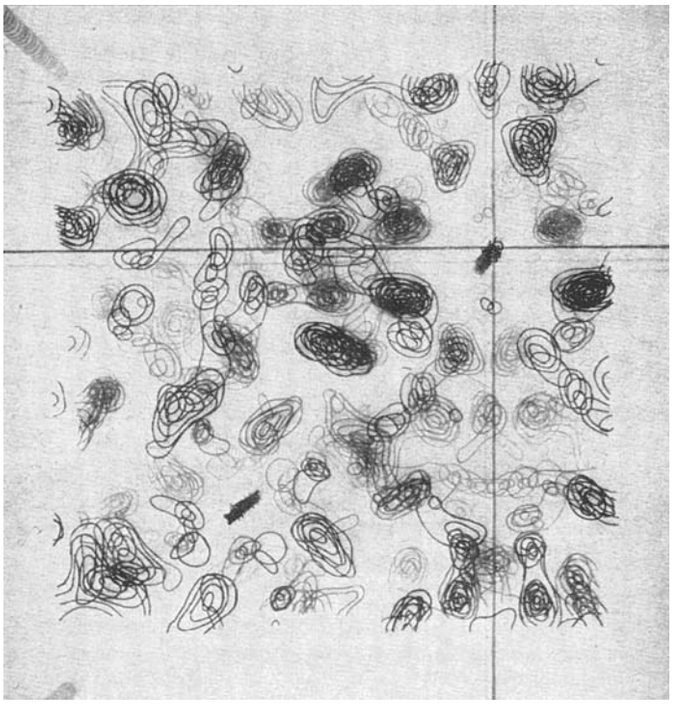

Fig. 26.1.2.12. Electron-density distribution in lysozyme at 6 Å resolution viewed parallel to the *c* axis. The horizontal and vertical lines represent the twofold rotation axes and intersect the twofold screw axis, upper right of centre. The fourfold screw axis is at the lower left of centre. The contour 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,

however, neither analysis yielded much information on the structures of these proteins, and protein crystallographers generally found them discouraging. Indeed, some went so far as to suggest that only the structures of proteins with a high α-helical content would be amenable to study by X-ray methods. Our reaction, however, as we turned our attention to extending the study of lysozyme to high resolution, was to resolve that each step in the analysis must be conducted as well as possible. We considered carefully, therefore, what improvements might be made to the methods employed hitherto. In particular, we recognized that further work would be needed to identify heavy-atom derivatives suitable for use at high resolution, and we sought improvements in the methods used for data collection, the correction of absorption errors and the use of anomalous scattering in phase determination.

At a purely practical level, one of our major concerns was that our limited access to the London University computer would lead to serious delays in the next stage of the work, which was bound to be even more dependent on computing than the work at 6 Å. Accordingly, we sought support from the Medical Research Council (MRC) for the acquisition of a laboratory-based computer that would be able to handle the computations up to but not including the calculation of a high-resolution electron-density map. Happily, the MRC provided a grant for an Elliott 803B computer, which was installed in the laboratory in March 1963. At the same time, the MRC provided a grant to purchase the commercial version of the linear diffractometer, which was manufactured by Hilger & Watts, Ltd. Since the Elliott 803B had not previously been used for crystallographic computing, this change in our computer involved many members of the laboratory in new programming.

26.1.3.1. *Heavy-atom derivatives at 2 Å resolution*

The potential usefulness of the three derivatives used at 6 Å resolution for phasing a higher-resolution map was analysed by CCFB and DFK. As the MHTS derivative had much the highest R factor at 6 Å resolution, and the K_2HgI_4 derivative had problems of stability and structure, only the K_2PdCl_4 (PD) derivative seemed likely to be useful for phasing at higher resolutions. An immediate search for additional heavy-atom derivatives was therefore undertaken, which included a re-examination of uranyl nitrate, $UO_2(NO_3)_2$ (UN). Together with other compounds, DFK obtained samples of the then novel $UO_2F_5^{3-}$ ion (UF) from Reuben Leberman at Cambridge, which generated a different pattern of changes in the lysozyme diffraction pattern to any of the previous heavy atoms. When the native phases from the 6 Å map were applied to the UF changes in the centric $hk0$ and $h0l$ zones, they showed a novel two-site binding pattern with a low R factor.

The UF and PD derivatives were examined at 2 Å resolution. Photographs of the centric $hk0$ and $h0l$ zones were taken with a 23° precession angle, the intensities were measured on the Joyce–Loebl densitometer and corrected for Lorentz and polarization effects by a program written by CCFB for the Elliott 803B computer. The heavy-atom parameters obtained in the refinement of these derivatives at 6 Å were used as a starting set for the refinement at higher resolutions. This refinement, like that at 6 Å resolution, was carried out with the program (*Pangloss*) based on Hart's (1961) method, but improved and rewritten for the Elliott 803B by DFK.

Initially only the PD and UF derivatives were refined: the mercuri-iodide derivative was not seriously considered to be a potentially useful derivative at high resolution because of the loss of heavy atom with irradiation observed during collection of the 6 Å data. However, it appeared probable that a mean occupancy would be suitable for photographic data, and that the signs predicted by the derivative might be very useful when the other two derivatives gave weak or ambiguous pre-

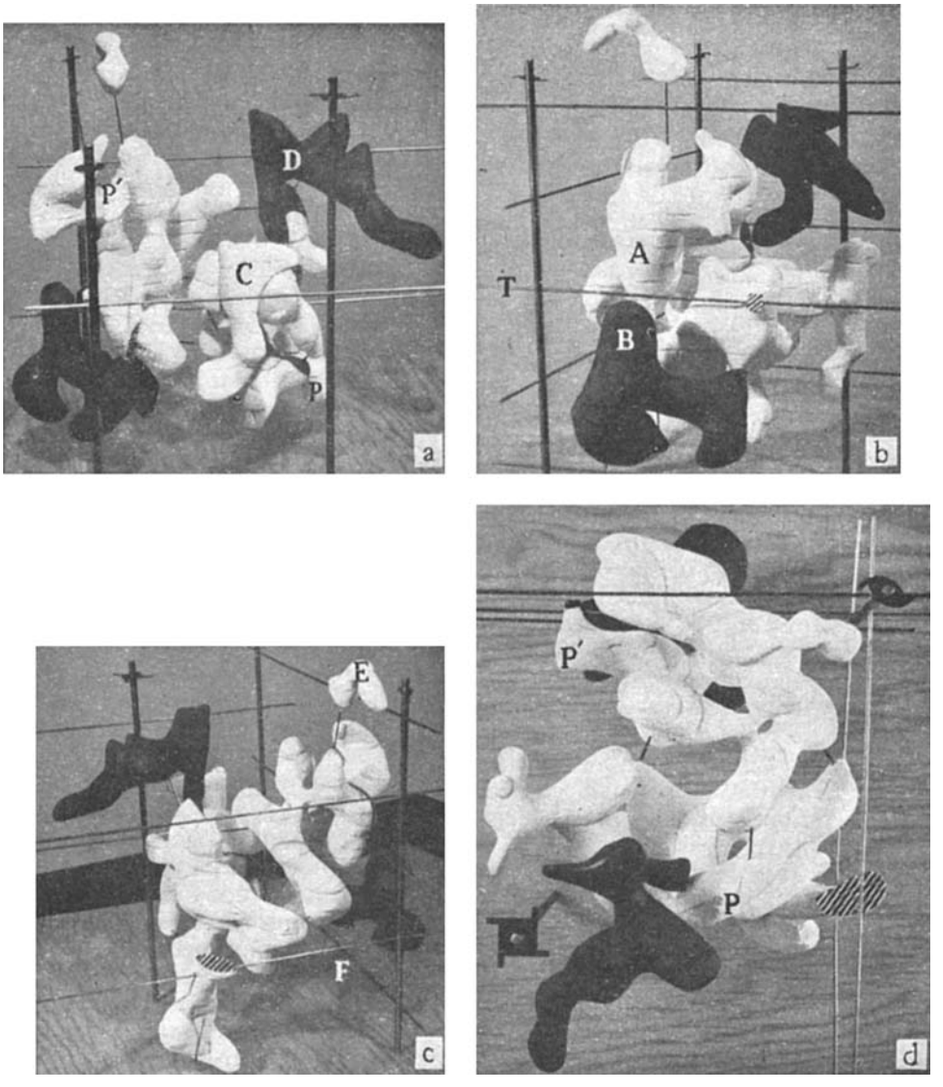

Fig. 26.1.2.13. Views of a 6 Å resolution model of the regions in which the electron density exceeds about 0.53 e Å^{-3}. The vertical rods indicate the twofold and fourfold screw axes parallel to **c**. The thinner horizontal rods are twofold rotation axes. At two places, shown hatched, rotation axes pass through continuous regions of density. The white regions alone make up the most compact asymmetric unit of the structure. In part (*b*), the grey section (B) is alternative to the region (A), to which it is related by the twofold axis T. In part (*a*), the grey section (D) is alternative to the white (C), related by the fourfold screw axis. The grey piece (F) in part (*c*) is alternative to the white (E), to which it is related by a unit-cell translation along **c**. The scale is indicated by the framework of symmetry elements, adjacent parallel twofold axes being 18.95 Å apart. Reproduced with permission from *Nature* (Blake *et al.*, 1962). Copyright (1962) Macmillan Magazines Limited.

dictions. In fact, the PD and UF derivatives on refinement at 2 Å resolution predicted sets of signs that disagreed to an unacceptable extent, with the PD derivative having a much higher R factor. Refinement of parameters defining the central mercury atom of the mercuri-iodide derivative, followed by calculation of a high-resolution difference-Fourier map using only the most clearly predicted signs, revealed a planar trigonal HgI_3^- ion, whose shape and size were similar to those found by Scouloudi (1965) and Fenn (1964).

Unfortunately, refinement of the mercury and iodine parameters produced a set of protein signs at variance with both previous sets. On the basis of the agreement between $F_H(\text{calc})$ and $F_{PH} - F_P$, both the PD and HgI_3^- derivatives appeared to be non-isomorphous with the native structure. The apparent inability of all three derivatives used at 6 Å to be of use in phasing reflections at higher resolution was disappointing. One possible explanation that was explored for the poor performance of the PD derivative at high resolution, when it showed large changes and a low R factor at 6 Å resolution, was the question of its structure. As with mercuri-iodide, the number of electrons in the halogen substituents exceeded the number in the central palladium atom. Much time was expended in trying to resolve the structure of the square-planar ion without success. This included examining all PdX_4 and PtX_4 complexes, where X is Cl, Br, or I, and calculating double-difference maps between $PdCl_4$ and $PdBr_4$, but without revealing any real indication of structure in the heavy-atom peak.

An extended search for alternative heavy-atom derivatives was therefore begun. Crystals for this purpose were grown as described above (Section 26.1.2.2), except that, on the initiative of DFK, small polyethylene bottles were used instead of glass vials. The polyethylene bottles did not affect crystal size or quality, but allowed crystals to be detached from their surfaces without the damage that occurred when glass vials were used. The search was carried out by the diffusion method. A small number of crystals were isolated together with a known volume of mother liquor, and a solution of the heavy-atom compound, usually at a concentration to make 5 mM in the final solution, was put into a dialysis bag and suspended in the liquid above the crystals. After one or two days, one of the crystals was mounted and an offset 5° screenless precession photograph of the $hk0$ zone was taken. (In these photographs there is no overlap of the zero and first upper levels.) If the approximately 30 unique reflections in this zone showed little or no change in intensities, the crystal was usually not examined further, while those that showed some changes were set aside until a higher-angle precession photograph could be taken. By the use of this technique, a rapid survey of potential derivatives could be made, and those heavy-atom compounds that did not bind to the protein were rapidly dismissed.

Crystals that showed substantial differences in intensity with respect to the native enzyme were used to take $hk0$ and $h0l$ precession photographs. In these studies, the precession angle was reduced to 15° in order to reduce the amount of data to be collected to more manageable proportions, but otherwise the data were collected as for the initial three derivatives. The heavy-atom parameters were also refined in the same way. However, to cope with the disagreements observed between the initial derivatives, a procedure was introduced to combine the refinement of individual derivatives with a refinement of the protein signs in the two centric zones, using all current derivatives together.

When a potential new heavy-atom derivative was identified, the procedure worked as follows. Difference-Fourier maps were calculated, initially at 6 Å and later, when most of the 2 Å signs had been established, at the maximum resolution of the derivative data. When a heavy-atom binding site had been located, its parameters were refined by Hart's method, and the consequences of the refinement, F_P and F_{PH} with their signs and the calculated F_H values, were compared with the corresponding results from other derivatives. The signs of each F_P predicted by the current derivatives were noted, and the most probable sign was determined for each reflection. Initially this was done by inspection, and later by the method of Blow & Crick (1959), with E values provided by the *Pangloss* refinement. The 'best' signs given by this procedure were used to calculate a double-difference map for each derivative with coefficients $[(F_{PH} - F_P) - F_H(\text{calc})]$. When these maps showed either new sites or some new feature at the sites previously included, appropriate alterations were made to the current model of that derivative and further refinement was carried out. If, on the other hand, a derivative showed a high background without interpretable features, it was omitted from further cycles and replaced by another derivative so that a group of four to six derivatives was always in use.

After each derivative had been examined in this way, the set of 'best' signs was updated with the new models of the derivatives, and the procedure was repeated. This procedure worked very efficiently, rapidly indicated a non-isomorphous derivative – or rather one that was significantly less isomorphous than the best – and clearly showed features such as new sites, structure around previously included sites, incorrect or anisotropic temperature factors, and incorrect positional parameters or occupancies. Finally, when the five best derivatives were left in the list, it was immediately apparent, from inspection of the sign predictions in comparison with the E values, which derivatives should be used to phase the high-resolution map.

This method confirmed that MHTS, one of the derivatives used in the 6 Å map but then discarded, was useful at 2 Å resolution and revealed that two new derivatives, UN and UF, were also satisfactory. The poor performance of the MHTS derivative at 6 Å resolution was due to anomalously large changes in the very low resolution reflections, which are sensitive to salt concentrations, but it exhibited good isomorphism at higher resolutions. At about this time, in the summer of 1963, DFK left the laboratory to take up an appointment at Brookhaven National Laboratory.

It became apparent during the course of the heavy-atom search and refinement that the number of suitable derivatives was severely restricted by a feature of the protein. This was the existence of a close pair of very strong binding sites, whose occupation was always accompanied by non-isomorphism. These sites, referred to by the initial derivatives that they bound as the HgI_3^- and PDsites, were found to have a strong affinity for all the complex halogen anions of Pd^{II}, Pt^{II}, Pt^{IV}, Au^{III}, Hg^{II}, Os^{IV}, Ir^{III} and Ir^{IV} that were tried. The sites were mutually exclusive, probably because they shared a protein side chain that acted as one of the important metal-binding groups. The HgI_3 site bound HgI_3^-, $PtCl_4^{2-}$, $PtBr_4^{2-}$, PdI_4^{2-}, $OsCl_6^{2-}$, $IrCl_6^{3-}$ and $AuCl_4^-$. All these compounds caused disordering of the protein structure, as indicated by a decrease in intensities at high resolution, and at least two of the compounds were sensitive to X-irradiation. The PD site also bound $HgCl_2$, $PdBr_4^{2-}$, $PtCl_6^{2-}$, $PtBr_6^{2-}$ and PtI_6^{2-}. All these derivatives were seriously non-isomorphous with the native structure at medium to high resolution, but careful analysis suggested that $PtCl_6^{2-}$ would be useful at low resolution. In contrast, the two uranium compounds, UN [which most probably gave rise to a bound $UO_2(OH)_n$ cation] and UF, gave a substitution pattern entirely different from other derivatives and, in particular, avoided the two sites that gave so much trouble with other complex ions. This finding is in accord with their known tendency to complex with protein oxygens, as opposed to the protein nitrogens that form the binding sites of most other heavy metals.

The two mercury benzene sulfonates that were investigated, MHTS and PCMBS, had a common sulfonate site, but the mercury atoms in the two derivatives were found to be about 3 Å apart. This suggested that they were bound to the protein by their charged

sulfonate groups. A similar orientation and position of the benzene ring was implied by the observations that the Hg–SO_3 distance was 5.76 Å in the MHTS derivative, compared with 5.67 Å in the crystal structure of MHTS (Fenn, 1964), and 7.00 Å in the PCMBS derivative, compared with 7.25 Å in the structure of PCMBS. Moreover, the angle Hg(MHTS)–SO_3–Hg(PCMBS) was found in the complexes with lysozyme to be 32°, the same as that calculated from the mercurial structures. This indicates that the benzene sulfonate groups of these two derivatives were fixed in the same position in the lysozyme crystals, and the location of the mercury in the protein crystal depended solely on its position on the benzene ring.

It is interesting to notice that PCMBS, which was investigated first, caused the *c* axis of lysozyme to lengthen by 1.5% and consequently could not be used at high resolution, while the slightly different MHTS reduced the lengthening to only 0.25% and could be used. This was an early example of an engineered isomorphous derivative. Nevertheless, PCMBS seemed potentially useful at low resolution.

A total of about 50 compounds were used to prepare derivatives (Blake, 1968). Many of these compounds were available commercially but some of the most important, including MHTS and $UO_2F_5^{3-}$, were not. MHTS was designed and synthesised by JWHO, as described above, while the initial sample of $UO_2F_5^{3-}$ was provided by Reuben Leberman, and further supplies were synthesized by CCFB.

In order to prove these new heavy atoms in three dimensions and also to eliminate any 'cyclic' effect due to the use of Fourier methods, 6 Å three-dimensional data were collected early in 1964 for the UF, UN, MHTS, PCMBS and $PtCl_6^{2-}$ derivatives. Two- and three-dimensional Patterson functions were calculated and interpreted afresh. These interpretations were wholly consistent with the results from the high-resolution projections, with the exception of the UN complex. At 3 Å, the *hk*0 and *h*0*l* projections showed only two sites, but at 6 Å there appeared to be three additional sites. This point was cleared up when refinement of the 2 Å data collected for phase determination showed that the three extra sites had very high temperature factors, which would result in these peaks being very low at 3 Å resolution. These five compounds were used to phase a second 6 Å electron-density map of the enzyme, as described below, in advance of the analysis at 2 Å, which was based on the only three derivatives that had been found to be reasonably isomorphous at high resolution: MHTS and the two uranyl derivatives, UF and UN.

The discovery of the two uranyl derivatives was not entirely accidental. In the preliminary study at 6 Å resolution, we had been impressed by the potential utility of anomalous scattering in phase determination and had noted the very high value of the imaginary component $(\Delta f'' = 16)$ of the anomalous scattering by uranium (Dauben & Templeton, 1955).

26.1.3.2. *Intensity measurements*

A further advance in diffractometry arose from the observation by DCP that more than one reflection can be measured at the same time in the flat-cone setting of a diffractometer (Phillips, 1964). In the flat-cone setting, the crystal axis is inclined to the X-ray beam as in the equi-inclination setting (Fig. 26.1.2.4*b*), but the motion of the counter is confined to the plane perpendicular to the crystal axis. This flat-cone plane of the reciprocal lattice is midway between the zero-level and equi-inclination levels. When measurements are made in the flat-cone setting from crystals rotating about a reciprocal-lattice axis perpendicular to reciprocal-lattice planes, the crystal and counter settings for reflections in levels adjacent to the flat-cone level are identical to one another and closely similar to those for reflections in the flat cone. This setting is illustrated in Fig. 26.1.3.1.

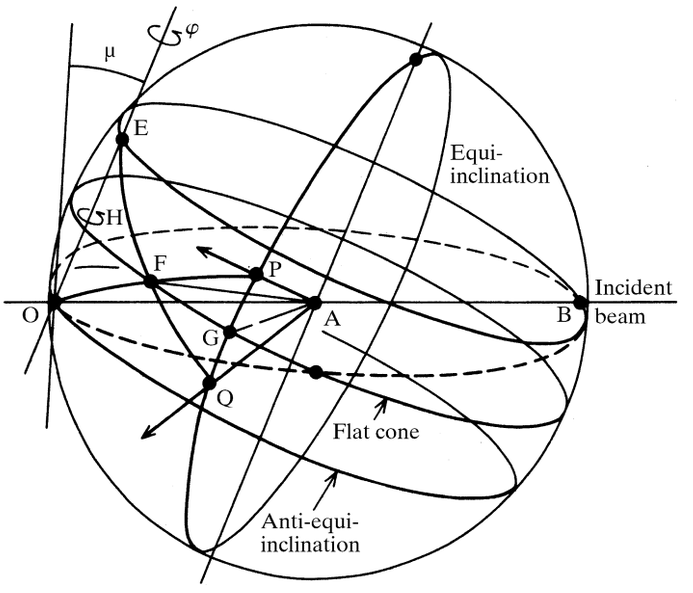

Fig. 26.1.3.1. Perspective drawing of the sphere of reflection showing inclination geometry with simultaneous reflections (reciprocal-lattice points P and Q) from levels symmetrically related by the flat-cone setting. Reproduced with permission from Phillips (1964). Copyright (1964) Institute of Physics.

This property of flat-cone geometry made it possible to modify the linear diffractometer so as to measure three reflections quasi-simultaneously (Arndt *et al.*, 1964). Even more reflections could be measured in this way from crystals with large unit cells but, at this stage, three 20th Century Electronics side-window proportional counters were mounted in echelon on the counter arm, with appropriate entrance slits and with their windows 0.75 cm apart. The distance of this array from the crystal could be varied between 22 and 50 cm so that the angular separation of adjacent counters lay between 0.034 and 0.015 rad. The separation of reciprocal-lattice levels in which quasi-simultaneous measurements could be made also varied, therefore, from 0.034 to 0.015 reciprocal-lattice units (r.l.u.'s), corresponding to crystal-lattice dimensions of 45 to 100 Å with copper $K\alpha$ radiation.

The array of proportional counters could be set in position on the counter arm by means of horizontal and vertical fine controls. The reflected X-rays passed through adjustable slits before entering the counter windows, and the left and right sides of these slits, or their top and bottom halves, could be blocked to facilitate precise setting.

A disadvantage of using flat-cone as opposed to equi-inclination geometry is that reciprocal-lattice levels measured in a flat cone, and adjacent levels, have a blind region at their centre in which reflections are not accessible. With the linear diffractometer, this blind region falls within the region in which the diffractometer does not automatically set the crystal and counter precisely enough for reliable measurements. The problem is not severe for measurements at low resolution, and it was avoided during the measurement of high-resolution data by the operation of low-angle limit switches that controlled the operation of the scanning and stepping slides. Fortunately, the high symmetry of the lysozyme crystals greatly reduced the seriousness of this problem, since reflections in the blind region close to the rotation axis could usually be measured as symmetry-equivalent reflections in the measurable region.

The first measurements made by this method were 6 Å data for the native protein and the five derivative crystals that had been identified as giving useful phase information at low resolution. Crystals were first mounted to rotate about their *a* axes, and the crystal-to-counter distance was 38.5 cm. Unfortunately, this long

crystal-to-counter distance gave rise to weak measured intensities, because of the significant absorption of the reflected X-rays by the air in the counter arm. For this reason we considered filling the counter arm with hydrogen, but came to the conclusion that a simpler approach would be to make the bulk of the measurements from crystals mounted to rotate about their $[\bar{1}10]$ axes.

In this setting, the pyramidal end of the crystals fitted against the wall of the capillary tube in which they were mounted (Fig. 26.1.2.5*b*). The roughly square cross section of the crystals perpendicular to the rotation axis tended to minimize the absorption variation, which was significant with crystals of linear dimensions of about 0.5 mm. The use of large crystals was necessary with a relatively weak X-ray source running at 800 W and with a foreshortened focal spot of 0.4 × 0.4 mm.

It was convenient to index the reflections in a monoclinic cell, as shown in Fig. 26.1.3.2. In this orientation, the reciprocal lattice presented a diamond pattern to the triple-counter array, whose windows were set parallel to the rotation axis. The reflections therefore occurred in two intersecting sets of levels, the odd and even levels, which had to be measured separately. However, the need to collect alternate levels in this way conferred the advantage that the counters could be positioned closer to the crystal (27.1 cm) than in the *a*-axis mounting so that $>70\%$ of the reflected X-rays were transmitted to the counters. Despite the complexity that this geometry introduced at the data processing and reduction stages, the significant advantages that it offered at the experimental stage ensured its use. The use of a c^* mounting would have been even more advantageous, but it was ruled out both by the difficulty of mounting the crystals in this orientation and by the fact that the counter arm could not be set to the required length of 18.5 cm.

Native data were collected both from crystals rotated about the *a* axis and from crystals rotated about $[\bar{1}10]$ so that they could be scaled together to form a consistent set of three-dimensional data.

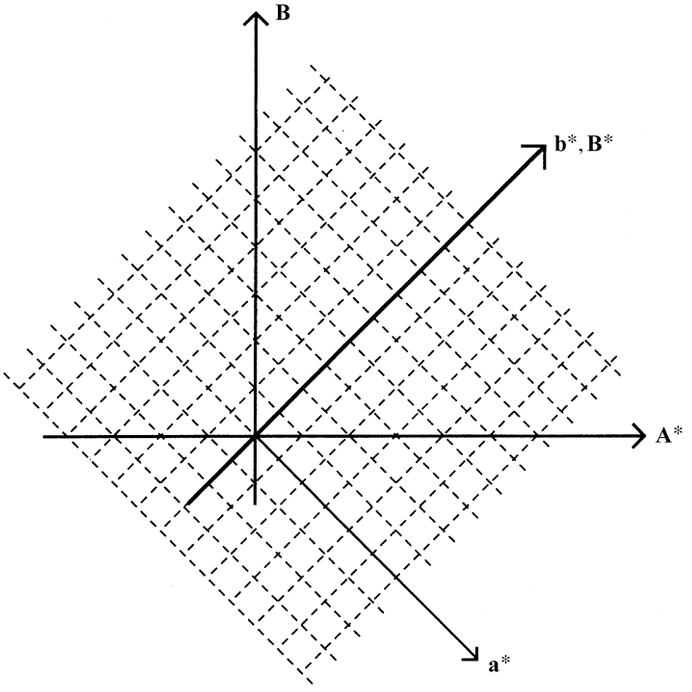

Fig. 26.1.3.2. The a^*b^* plane of the reciprocal lattice oriented to rotate about the $[\bar{1}10]$ axis. The indexing of reflections in a monoclinic unit cell with axes A^* and B^* is also shown. A^* coincides with [110] in the tetragonal lattice and B^* coincides with b^*. The rotation axis is the *B* axis in the monoclinic cell. The reciprocal-lattice dimensions with Cu $K\alpha$ radiation are $a^* = b^* = 0.0195$; $c^* = 0.0406$ r.l.u., while the cell diagonal in the a^*b^* plane is 0.0276 r.l.u.

For the derivatives, however, data were collected only from crystals rotating about their $[\bar{1}10]$ axes and these data were scaled together to form two sets, one comprising levels with odd indices corresponding to the rotation axis and the other with even indices. These measurements, comprising some 1200 reflections, could be made from a single crystal exposed to the X-ray beam for about 20 h. The odd and even data sets for the derivatives were then scaled separately to the native data to give complete sets of three-dimensional data. These low-resolution measurements were made during the first half of 1964 and, after processing by the methods described below, they were used in September 1964 to calculate a new image of the structure at 6 Å resolution.

26.1.3.3. *The second low-resolution map at 6 Å*

Our purpose in calculating a new electron-density map at 6 Å was fourfold. First to ascertain whether the procedures used to identify the five derivatives thought to be satisfactory at this level of resolution had worked satisfactorily. Second, to judge the quality of the measurements made by the triple-counter diffractometer. Third, to explore the effects of the modified method of applying absorption corrections to the intensities that are described below, although these were not expected to have a very great effect at low resolution. Fourth, to examine the effectiveness of the new procedure for incorporating anomalous-scattering information in the phase determination, which is also described below.

Comparison of the two sets of structure amplitudes gave a conventional *R* value of 0.075, which is not particularly good – perhaps because of the comparatively large background values associated with these low-angle measurements. However, the mean figure-of-merit obtained in the new phase calculations was 0.97 as compared with the 0.86 obtained originally. The root-mean-square difference in electron density between the two maps was 0.012 e Å^{-3}, from which it may be judged that the two maps were very similar. Nevertheless, the outline of the molecule was certainly clearer in the new map, and within the molecule there was improved continuity, suggesting the course of a folded polypeptide chain, and there were a number of stronger rod-like features suggestive of α-helices. Two of these were prominent, running upwards from right to left, in the view of the new model shown in Fig. 26.1.3.3.

The result was very encouraging, and we therefore went ahead immediately with data collection at 2 Å resolution, using essentially the same methods. At the same time, we began to plan low-resolution studies of inhibitor binding to lysozyme, from which we hoped to derive information about the nature of the enzyme–substrate complex.

26.1.3.4. *Intensity measurements at high resolution*

At high resolution, measurements were made from only six levels from native crystals set to rotate about the *a* axis while a complete set of data was collected from crystals rotating about $[\bar{1}11]$. As at low resolution, the *a*-axis levels were important for scaling together the odd and even sets of levels measured from crystals rotated about the tetragonal $[\bar{1}11]$ axis. No high-resolution measurements were made from derivative crystals rotated about the *a* axis, the intention being to scale the odd and even subsets of derivative data directly to the native.

26.1.3.4.1. *Experimental methods*

The crystals can be thought of as roughly flattened on $(00\bar{1})$ with liquid between $(00\bar{1})$ and the capillary tube in which they were mounted. In order to minimize the effect of the liquid on absorption, we therefore measured the reciprocal-lattice hemisphere with *C* and c^* positive. Reflections were scanned along *A*, and the origin of the

scanning slide was offset for upper levels so that, for example, the centre of the scanning slide was the $H = -4$ for the level $K = 8$.

Crystals could be set fairly well by eye. Viewed through the diffractometer microscope, they looked roughly square in the [001] direction, and the arcs of the goniometer head were set so that the edges were horizontal and vertical. When the crystal was turned through 90° from this position, the reflection of a light held level with the microscope could often be seen in the true (110) face. Setting the crystals on the goniometer head with [001] parallel to one of the arcs facilitated subsequent adjustments.

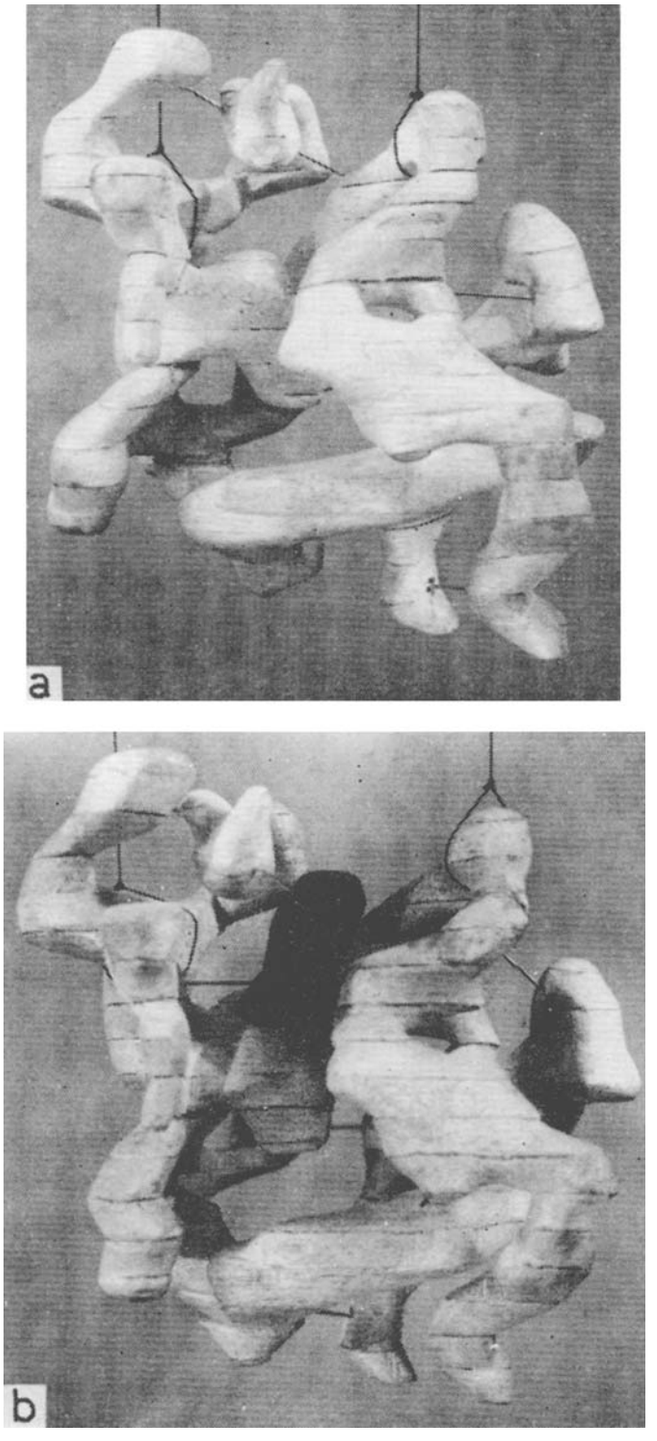

Fig. 26.1.3.3. Solid model of the electron density greater than about 0.5 e Å^{-3} in the second study of lysozyme at 6 Å resolution. This view of the model is equivalent to a view of the original model seen horizontally from the right of Fig. 26.1.2.13(*c*). (*a*) The new model has a marked cleft running roughly vertically down the other side of the model, corresponding to the one that can be seen in Fig. 26.1.2.13(*c*). (*b*) The cleft was shown to bind inhibitor molecules. The black density is that observed for the lysozyme–GlcNAc complex at 6 Å resolution.

Fine adjustment of crystal orientation was achieved by setting on the $\bar{4}40$ reflection. For this purpose, C – the vertical slide – was set to 0.1104 and the inclination angle to $\mu = 3° 5'$. Then, if one arc was set fairly well by eye, the other could be oriented parallel to the incident X-ray beam and adjusted to locate the reflection. The two arcs were then adjusted in turn to give the optimum setting, in which the crystal was rotated about the normal to the $(\bar{4}40)$ plane.

The orientation in the AC reciprocal-lattice plane was determined by returning the vertical slide to zero and setting the upper (scanning) slide to 0.1104 with the lower (stepping) slide at zero. The crystal was then rotated until the monoclinic 400 reflection appeared. At this stage, checks were made with the top/bottom and left/right slits to make sure that the crystal was well centred in the X-ray beam and that the counter apertures were well positioned. Similar checks were made with the crystal rotated about 180° to the monoclinic $\bar{4}00$ reflection.

The final check on crystal orientation was to locate the 008 reflection near 0.3248 on the lower (stepping) slide, with the other slides set at zero. There was some variation in the value of c^* for different crystals, and c^* was often closer to 0.0404 than to 0.0406 r.l.u.

At this stage, the first measurements were made of the intensities of the reflections (monoclinic) 400, 16,0,0 and 008. These reference reflections were remeasured at intervals during the measurement of each triplet of reciprocal-lattice levels as a check on the stability of the whole system and irradiation damage to the crystal. The measurements were manually entered on the diffractometer output tape and monitored by the data-processing program.

In order to set the diffractometer for a particular triplet of levels, we found it convenient to leave the lower slide set for the 008 reflection, to adjust the vertical slide and the tilt angle (μ) for the levels in question, and then to run out along the stepping slide until a suitable high-angle reflection was found. The crystal rotation angle was then optimized for this reflection.

Finally, careful checks were made to ensure that the reflections to be measured fell in the reciprocal-lattice hemisphere with L positive, and that they were indexed in a right-handed axial system. The automatic run was then begun at the 2 Å limit on the stepping (C^*) slide, and all reflections were scanned from $H_{\max}$ to $-H_{\max}$ on the scanning slide. Virtually all reflections within (and in some directions a little beyond) the 2 Å limit were measured in this way in fourteen triplet levels, seven even and seven odd. About 2000 reflections were measured in a typical overnight run, and data collection for each species of crystal took slightly more than two weeks.

The derivative crystals required for these measurements were prepared by the diffusion method described above, and about 20 of them were mounted at the same time at the beginning of the data-collection process to ensure that they were all the same. At the end of a run, careful measurements were made of the peak intensity of the $\bar{4}40$ reflection as the crystal was rotated through steps of 15° about the crystal-mounting axis (φ). These measurements were used during data processing in an improved method of absorption correction devised by North *et al.* (1968).

26.1.3.4.2. *Diffractometer output*

The output from the triple-counter diffractometer again consisted of a plain-language output for immediate checking of the results and output at this stage on eight-hole-punched paper tape for immediate

```
CRYSTAL PDBR 4  D 2 * 1/5  LEVELS  1 2:1 4:1 6  2 1/2/6 4

*0 1+1 4+0 1+
:0 0 0 3 4 6   0 0 0 3 3 6   0 0 0 3 2 8
:0 0 7 1 0 8   0 0 0 7 4 8   0 0 0 8 9 0
:0 0 0 3 7 8   0 0 0 4 0 8   0 0 0 3 0 3

*0 0−1 4+0 1+
:0 0 0 3 3 (   0 0 0 3 4 7   0 0 0 3 0 0
:0 0 0 7 8 5   0 0 2 0 5 5   0 0 0 6 5 2
:0 0 0 2 9 0   0 0 0 3 6 7   0 0 0 3 4 7

*0 1−1 4+0 1+
:0 0 0 2 9 6   0 0 0 3 4 9   0 0 0 3 5 6
:0 0 0 6 1 1   0 0 0 7 1 6   0 0 7 5 4 3
:0 0 0 3 1 2   0 0 0 3 4 3   0 0 0 3 2 2

*0 2−1 4+0 1+
:0 0 0 3 0 2   0 0 0 3 1 9   0 0 0 2 4 4
:0 0 5 8 9 9   0 0 2 3 9 8   0 0 1 5 7 0
:0 0 0 2 8 5   0 0 0 3 0 3   0 0 0 2 4 1

*0 3−1 4+0 1+
:0 0 0 3 7 7   0 0 0 3 6 9   0 0 0 2 7 3
:0 0 8 3 4 1   0 0 6 4 3 4   0 0 3 0 7 0
:0 0 0 3 0 2   0 0 0 3 3 9   0 0 0 2 3 0

*0 4−1 4+0 1+
:0 0 0 8 4 1   0 0 0 2 6 6   0 0 0 3 7 0
:0 1 4 9 1 6   0 0 0 9 7 5   0 0 0 8 4 4
:0 0 0 3 1 1   0 0 0 2 8 0   0 0 0 1 9 9
```

Fig. 26.1.3.4. Paper-tape output from the triple-counter linear diffractometer, showing the indices of the central reflection and the background, peak, background counts for the three reflections. Reproduced with permission from Arndt *et al.* (1964). Copyright (1964) Institute of Physics.

transfer to the computer. An example of this record is shown in Fig. 26.1.3.4.

26.1.3.5. *Data processing*

Although the in-house Elliott 803B computer provided much improved facilities over those previously available, its limited store capacity (4096 words of magnetic core memory, but no magnetic drum, disk, or tape facilities) resulted in data processing being carried out in a series of stages, each of limited complexity. Input and output was by means of eight-hole paper tape, which was used as the medium for intermediate storage. The output tape from each stage was used for input to the next stage, together with appropriate parameters. Manual input was restricted to such parameters as unit-cell dimensions and ordinates of the absorption curves, with all other input being in the form of computer-generated output from a previous stage, starting with the diffractometer output tapes. This approach resulted in a great reduction in manual labour and intervention compared with the low-resolution work on lysozyme and the high-resolution stage of myoglobin, and it was probably the first crystal-structure determination that was fully computerized.

As for the initial low-resolution work, the first stage of data processing was input and checking of the diffractometer output, with the programs modified appropriately to deal with the sets of three reflections measured simultaneously. The background-corrected intensities were then output together with a table of the reference reflections used to estimate the extent of any radiation damage. Corrections were then applied for radiation damage if required, followed by application of Lorentz–polarization corrections, followed by absorption corrections.

26.1.3.5.1. *Absorption corrections*

Although absorption of X-rays by protein crystals is low compared with crystals having a preponderance of heavier atoms, corrections for absorption are required in order to give F values that are sufficiently precise for calculation of the relatively small changes due to the introduction of heavy atoms or anomalous dispersion. The mounting of protein crystals within a glass capillary, normally with a small amount of mother liquor between the crystal and the capillary wall, presents a complicated situation for absorption calculations. Although Wells (1960) wrote a computer program to deal with the situation, a severe impediment to the use of theoretical methods of correcting for absorption results from the very great difficulty in obtaining precise measurements of the mounted crystal, the liquid meniscus and the capillary-tube walls. An alternative approach was to use a semi-empirical method and, for low-resolution lysozyme studies, Furnas' (1957) method had been employed, as described above.

Despite the fact that the method of Furnas had been successful in improving the agreement between symmetry-related reflections in the earlier studies, the implicit assumption that the absorption depended upon the mean direction of the incident and reflected X-ray beams became clearly less valid as the Bragg angle increased. We therefore implemented a development of the method in which the absorption correction applied to any reflection was given by the mean of the two values for the directions of the incident and reflected beams (North *et al.*, 1968).

Although the method was easy to apply and was of significant value, as judged by the improved agreement between symmetry-related reflections, it nevertheless provided only a partial correction for absorption, because of the assumption that absorption is dependent solely on the directions of the incident and reflected beams. The limitations of this assumption are particularly important where precise values are required for Friedel pairs of reflections in order to make use of anomalous-scattering differences in phase determination. Fig. 26.1.3.5 shows two contrasting situations that can arise when the environment of a crystal is asymmetrical because of its mounting. The Friedel pair of reflections shown in Fig. 26.1.3.5(*a*) would suffer similar absorptions, whereas the pair shown in Fig. 26.1.3.5(*b*) would have significant differences because of the location of the mother liquor. In the 2 Å structure determination of lysozyme, an approximate correction was made for this effect. From Fig. 26.1.3.5, it is clear that the absorption error arising from the asymmetric distribution of the mother liquor is 0 for reflections with $h = 0$ and becomes increasingly great as h increases. The assumption was made, therefore, that the required correction was a function only of h and the reflections hkL with constant L were divided into groups with constant h and $-h$. The ratios $\sum_k I(h)/\sum_k I(-h)$ were then plotted against h, as shown in Fig. 26.1.3.6.

Such plots were frequently found to be linear, and the corresponding linear correction was then applied to each row on the more highly absorbed side in order to bring its mean intensity up to that of the other. Where the plots were not linear, so that a simple form of correction was not applicable, the entire set of measurements was usually rejected.

26.1.3.6. *Further stages of data processing*

Many of the programs for subsequent stages of data processing were written by VRS, who had discussed the work with DCP at a meeting in Madras in January 1963 and who joined the team in October 1963.

As described in Section 26.1.3.2, the native data comprised three sets, one of six reciprocal-lattice levels measured from crystals rotated about the a axis, and the other two comprising the odd and even subsets of levels collected about the $[\bar{1}11]$ axes. Within these three sets, the odd and even $[\bar{1}11]$ levels had no rows in common with one another, but each had rows in common with the a-axis data. Extraction of the related rows permitted the calculation of scale factors by the method of Hamilton, Rollett & Sparks

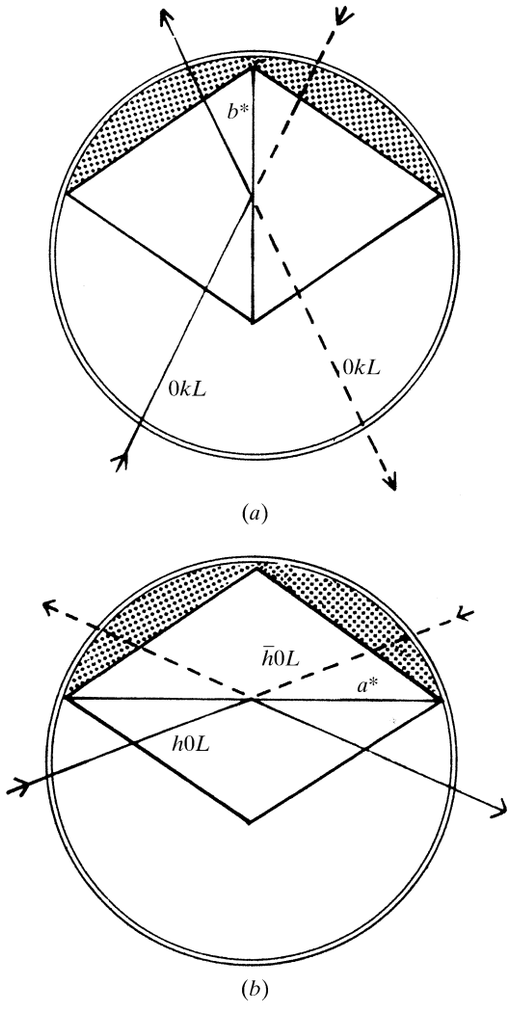

Fig. 26.1.3.5. Asymmetric mounting of a protein crystal with its mother liquor in a capillary tube. Anomalous-scattering differences would be seriously affected in (*b*) but not in (*a*). Reproduced with permission from North *et al.* (1968). Copyright (1968) International Union of Crystallography.

(Hamilton *et al.*, 1965). Application of these scale factors produced a complete set of self-consistent native data.

The derivative data, which consisted of the odd and even levels collected about the [$\bar{1}$11] crystal axes, were scaled to the native data by the application of two scale factors derived by comparing the totals of the intensities in the odd and even subsets. Because of the high symmetry of the tetragonal space group, up to eight measurements were available for many symmetry-equivalent reflections; for the heavy-atom derivative crystals, these formed four sets of Friedel-related pairs. At this stage, the X-ray amplitude data were on 40 paper tapes for each of the native and three derivative crystals, one for each reciprocal-lattice level. The native and derivative data were then compared level by level, but it was found that the ratio was very nearly a constant, independent of sin θ. A further numerical factor was applied to all data to bring them approximately to an absolute scale. In the final stage of data processing, the data from the native and the three derivative sets were brought together into a single list containing the native *F*'s and the Friedel pairs for each of the three derivatives. In addition, all of the centric reflections were extracted in order to provide the data for refinement of the heavy-atom parameters.

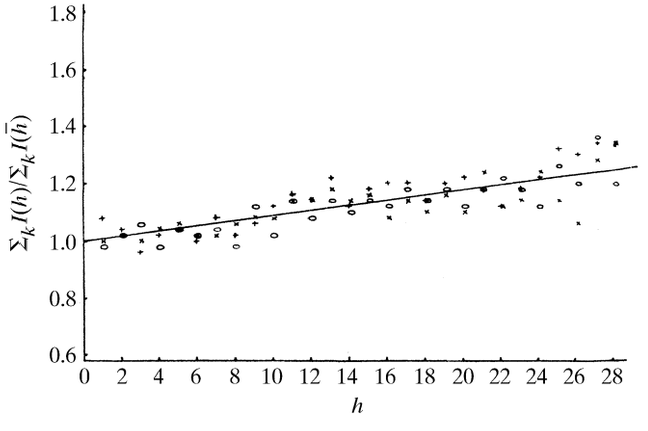

Fig. 26.1.3.6. Plot of the ratio $\sum_h I(h)/\sum_h I(-h)$ against *h*. In this example, a linear correction may be safely applied to equalize the average intensities in opposite rows (North *et al.*, 1968).

At this late stage in the data processing, we noticed that some levels agreed significantly less well than the norm with intersecting levels. On 23 October 1963, we had to recognize that the high-resolution measurements had been made from two different types of crystals with essentially identical unit-cell dimensions, but with subtly different diffraction patterns. We designated these two crystal types type I and type II, and set about the task of producing sets of data of one type to use in the structure analysis.

26.1.3.7. *The crystal-type problem*

This discovery was particularly galling because, although at the time variations in diffraction patterns had been reported for some protein crystals, we had failed to notice that this phenomenon had been mentioned years earlier in a study of lysozyme (Corey *et al.*, 1952).

A preliminary analysis of the differences in the diffraction patterns suggested two important characteristics:

(1) the differences tended to increase with resolution and

(2) the differences appeared to be more consistent with two discrete diffraction patterns than a continuum of patterns lying between two extremes.

The two diffraction patterns were characterized operationally by specific patterns of intensities in the 3–4 Å resolution range (where the differences appeared to be maximal), and particularly by a few adjacent pairs of reflections whose relative intensities interchanged in the two types. The principal diagnostic reflections were 11,11,4 and 11,11,5. In data associated with crystal type I, $I(11,11,4) > I(11,11,5)$, while in crystal type II, $I(11,11,4) < I(11,11,5)$. The 11,11,*L* rows of reflections from the two types of crystals had the structure amplitudes shown in Table 26.1.3.1.

In order to stand any chance of successfully calculating the lysozyme map by isomorphous replacement, we had to sort all the data so far collected, both native and derivative, into the two types and then recollect 'rogue' data sets in order to assemble complete data sets of one particular type. The alternative was to recollect the whole data on a sounder basis, which we were loath to do, especially as other teams seemed likely to be well advanced in their solution of the lysozyme structure.

It appeared that the bulk of the data that had already been collected using the diffractometer was what we had called type II. We also observed that nearly all the photographic data collected in the heavy-atom proving stage was of type I. This observation was of great importance because it gave us a sound basis for defining the differences between the type I and II diffraction patterns, and it also provided a vital clue in identifying from their shapes the crystals that gave the two types of diffraction pattern. This was very important to us in the selection of crystals to replace the rogue data

Table 26.1.3.1. *Structure amplitudes of the* 11,11,*L reflections from crystal types I and II*

Crystal type	Reflection					
	11,11,0	11,11,1	11,11,2	11,11,3	11,11,4	11,11,5
I	29.2	21.8	10.6	18.2	23.1	17.0
II	36.9	25.7	19.7	27.7	26.4	36.8

sets. The crystals that gave the best results for the photographic work tended to be relatively small and flattened along the tetragonal fourfold axis direction, while those that gave the best diffractometer data were the larger isometric crystals, which were more extended along the crystal fourfold axis (Fig. 26.1.2.1). These crystals could be definitely associated with the type I and type II diffraction patterns, respectively. Batches of lysozyme crystals grown according to the procedures defined earlier usually contained both types of crystal. This suggested that the two crystal forms might have originated when the pH of crystallization was on the borderline between two crystal forms of lysozyme. This hypothesis was supported by the observation that crystals grown at a somewhat higher pH had diffraction patterns closely similar if not identical to type I.

The more thorough analysis of the differences between the diffraction patterns for types I and II that these findings permitted showed that the differences in the intensities of equivalent reflections were resolution-dependent. They were very small in the 6 Å region (which probably accounts for the differences not being observed earlier), increased to a maximum at the position of the normal 4 Å peak in the protein diffraction pattern, and fell off at higher resolutions. This pattern is consistent with a lack of isomorphism between the two types of crystal of the kind that may be caused by a slightly different orientation of the lysozyme molecule in the unit cells of the two crystals (Crick & Magdoff, 1956). Such effects may be brought about, for example, by slightly different charge distributions in the protein molecules, rather than by the presence of additional diffracting material in one crystal type or the other. [This conclusion was later confirmed by a detailed analysis of the two structures by Helen Handoll (1985).]

These observations suggested it was safe to go ahead with trying to determine the structure of lysozyme in either type of crystal, and the decision to proceed with the type II crystals was solely on the basis that the bulk of the data already collected were of this type. Knowing the characteristic pairs of reflections that distinguished the two crystal types, we found it relatively straightforward to ascribe each data set to a particular crystal type, even for the heavy-atom derivatives, because the type differences were much larger than the heavy-atom changes at approximately 4 Å resolution. All triplet levels of data that belonged to the type I diffraction pattern were extracted from the total data sets and replaced by equivalent data sets recollected from confirmed type II crystals. This process, whose success was carefully tested and confirmed during the data processing and reduction stage, resulted in consistent data sets at 2 Å resolution for the native lysozyme and all three isomorphous derivatives that were derived from type II crystals.

26.1.3.8. *Final refinement of heavy-atom parameters*

At this stage, we were able to review the refinement of the heavy-atom parameters in relation to the type II crystal data that were now available. Prompted in part by the observation that the heavy-atom positions in myoglobin appeared in regions of high negative density in the final electron-density map, we considered the need to adopt an improved way of modelling the heavy atoms, particularly for the complex uranium compounds, in which the uranium atoms were coordinated to several oxygen or nitrogen atoms. Calculation of the R factors and occupancies of the heavy atoms as a function of $\sin\theta$ showed that the heavy-atom compounds, particularly the uranyl complexes, could not be modelled well by a single atom, but that it was necessary to take into account both the O and the N atoms in the complex and the water that had been displaced by the heavy-atom cluster. The difference electron-density maps did not show the orientation of the O and N atoms, and our first thought was to model the complex in terms of a central U atom, surrounded by a spherical shell of electron density representing the O and N atoms minus a sphere of electron density representing the displaced water. Trial calculations suggested that this model would give improved agreement between observed and calculated $|\Delta F|$ values, most significantly for the uranyl derivatives. On second thought, it occurred to us that a more satisfactory approach might be to derive empirical scattering-factor curves by fitting the curves representing the variation of heavy-atom occupancies with $\sin\theta$ to polynomial functions of $\sin^2\theta$. This was easily accomplished by the use of a standard curve-fitting program and, for ease of use in the phase program, these scattering curves were fitted to curves of the type $f(H) = a + b\sin^2\theta + c\sin^3\theta + d\sin^4\theta$. With this modification, *Pangloss* was used with the type II crystal centric data to obtain the heavy-atom parameters shown in Table 26.1.3.2.

Table 26.1.3.2. *Heavy-atom parameters for the 2 Å structure*

O is the occupancy of the heavy-atom site (electrons); *B* is the isotropic temperature-factor constant; *E* is the root-mean-square difference between observed and calculated heavy-atom differences for centric reflections (electrons); *N* is the number of centric reflections in the range $0.01 < \sin^2\theta < 0.15$ used in the refinement and *R* is the reliability index for observed and calculated heavy-atom changes of centric reflections.

Derivative	Site	*x*	*y*	*z*	*O*	*B* (Å^2)	*E*	*N*	*R*
MHTS	I	0.2068	0.6138	0.0507	39.2	17.8	58	1247	0.60
	II	0.2415	0.6393	0.9326	8.8	14.9			
UF	I	0.1783	0.5849	0.7204	55.5	21.0	74	1277	0.52
	II	0.0974	0.8976	0.4650	29.3	24.3			
UN	I	0.0961	0.8938	0.2664	47.1	19.2	80	1140	0.57
	II	0.1898	0.5901	0.7168	42.1	124.8			
	III	0.0446	0.7266	0.5150	9.0	190.2			
	IV	0.0869	0.8976	0.4866	11.4	68.4			
	V	0.2024	0.6388	0.6781	28.6	42.8			

The temperature factors shown in Table 26.1.3.2 most deserve comment. All those obtained for the significant sites (of which there are no more than two for any derivative) are comparable with the overall value for the protein crystals themselves. The very large values obtained for other sites show that these sites are of little importance at high angles and may not represent real sites of heavy-atom attachment. The minor site of MHTS is clearly the SO_3^- group of this molecule (see Section 26.1.3.1).

26.1.3.9. *Calculation of phase values*

Blow (1958), in his determination of the phase angles of the noncentrosymmetric [100] zone of horse haemoglobin, and, later, Cullis *et al.* (1961, 1962), in their determination of the three-dimensional structure of horse haemoglobin, used anomalous-scattering data to supplement the information available from the isomorphous-replacement differences. In each of these studies, phase determination had been carried out by constructing probability curves from the multiple-isomorphous-replacement data and, when the most probable phase angle had been deduced, the anomalous-scattering data were examined. For many reflections, they allowed a choice to be made between two apparently equally probable values of phase angle given by the isomorphous data. This procedure was clearly rather arbitrary and subjective, and a method of combining anomalous scattering with isomorphous replacement in a more rigorous way was described by Blow & Rossmann (1961). In their method, which was subsequently employed for the low-resolution work on lysozyme, use was made of the fact that the mirror image of the Argand diagram for a $\bar{h}\bar{k}\bar{l}$ reflection is similar to the Argand diagram for the *hkl* reflection, but for the reversal of the sense of the imaginary part of the heavy-atom contribution. The data for the $\bar{h}\bar{k}\bar{l}$ reflections may therefore be treated as though they came from a separate isomorphous compound, with parameters identical to those of the original compound, but with the opposite sign for the imaginary component of the atomic scattering factor.

In the low-resolution lysozyme phase determination (Section 26.1.2.7), intensities of the Friedel pairs of reflections were measured for each of the three heavy-atom compounds, and the problem was treated as if there had been a total of six heavy-atom compounds. Although the method had been found helpful to some extent, analysis of the phases showed that the anomalous-scattering data had played comparatively little part in determining the positions of the centroids of the phase probability distributions, even for reflections with apparently significant anomalous differences.

ACTN observed that this apparent contradiction is because of the fact that the anomalous differences between Bijvoet pairs of reflections measured from the same crystal are inherently more accurate than the isomorphous differences that are measured from different crystals and subject to different systematic errors (North, 1965; Phillips, 1966). Indeed, analysis of the equivalent reflections from native and derivative crystals (Section 26.1.3.6) showed that the r.m.s. error E', corresponding to the anomalous differences, was about one-third of E, the error in the isomorphous differences. The result of incorporating this distinction in the phase program is illustrated in Fig. 26.1.3.7. Phase calculations for the new 6 Å and 2 Å maps of lysozyme were therefore carried out by using ACTN's method, with E' set at one third of E.

The data tapes containing the F values for the native and the Friedel pairs of F values for the three derivatives were used as input to a phase program written by ACTN. For acentric reflections, phase probabilities were calculated as described in the previous section, and the centroids of the distributions were determined in order to derive a 'figure of merit', which was applied to the structure amplitudes, as first proposed by Blow & Crick (1959), so as to produce a 'best' Fourier map. For the quite high proportion of centric reflections in the lysozyme diffraction pattern, phase probabilities were calculated by the formula appropriate to the case in which the native and derivative F's are collinear with each other and with the vector due to the heavy atom.

The phases of the 9040 reflections were calculated on the Elliott 803B computer and had a mean figure of merit of 0.60. The variation with angle was very similar to that obtained with sperm-whale myoglobin and is shown for the centric and acentric reflections separately in Fig. 26.1.3.8.

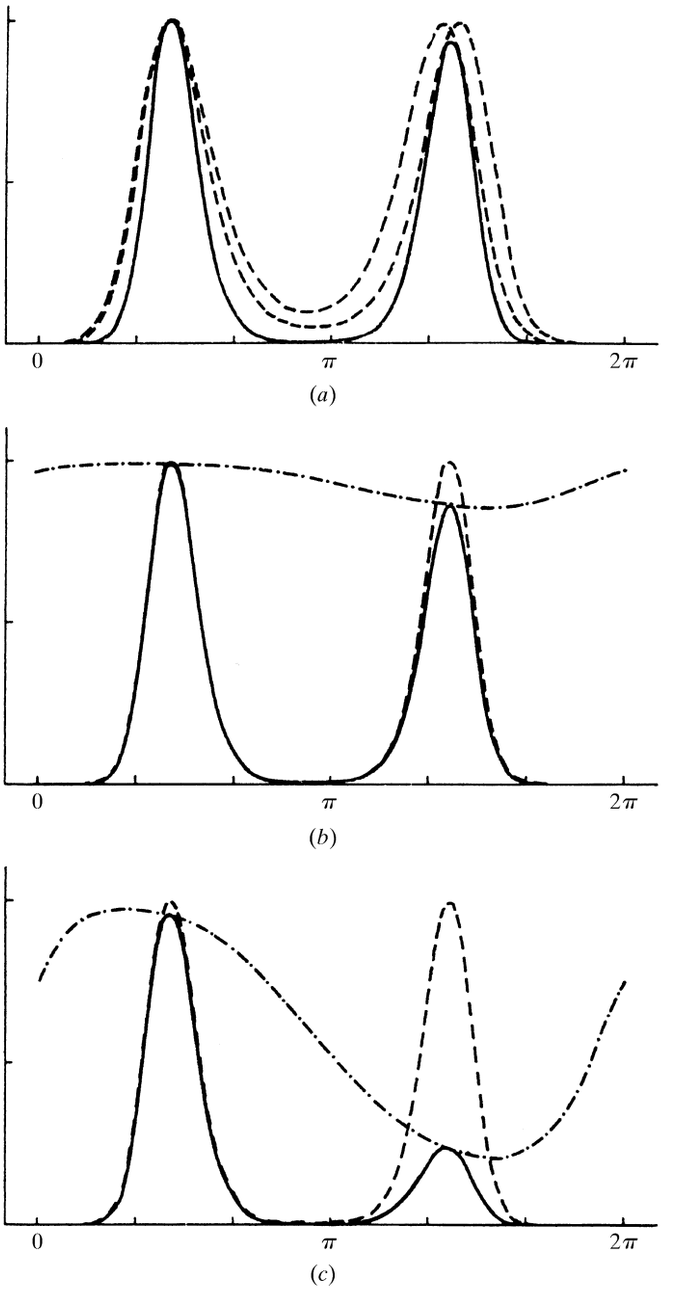

Fig. 26.1.3.7. (*a*) Phase probability curve for a Bijvoet pair of reflections (broken lines) with the joint probability curve (full line) derived by the method of Blow & Rossmann (1961). (*b*) Isomorphous-replacement phase probability curve derived from the mean of F_{PH^+} and F_{PH^-} (broken line); anomalous-scattering probability curve (chain line); and joint probability curve (full line) derived by the method of North (1965), using $E'(\text{anomalous}) = E(\text{isomorphous})$. (*c*) As (*b*), but with $E' = (1/3)E$. Reproduced with permission from North (1965). Copyright (1965) International Union of Crystallography.

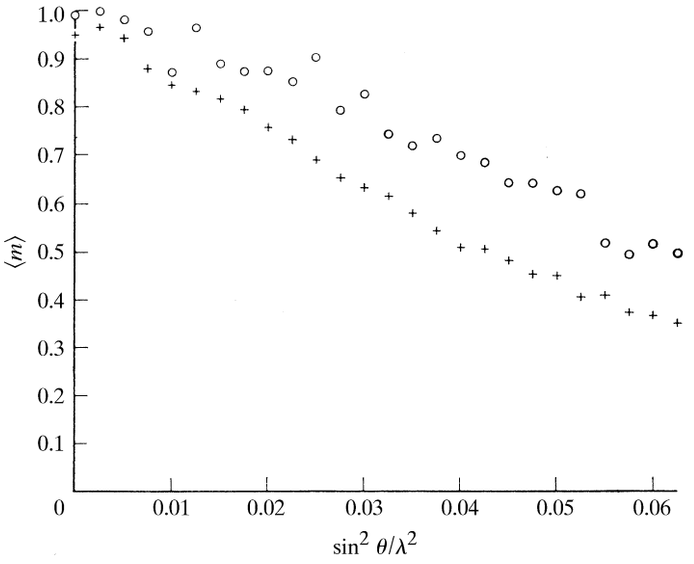

Fig. 26.1.3.8. Variation of the mean figure of merit $\langle m \rangle$ with $\sin^2\theta/\lambda^2$. (crosses represent acentric reflections, open circles represent centric reflections).

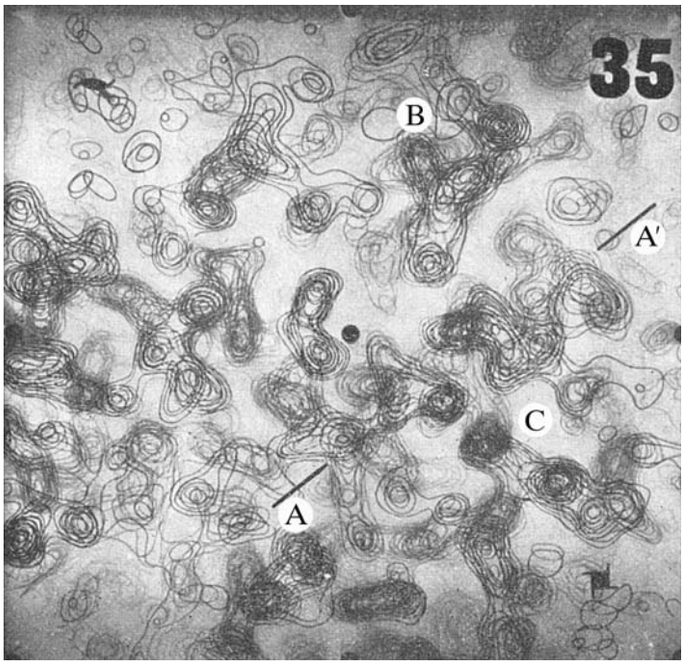

Fig. 26.1.3.9. Photograph of sections $z = 35/60$ to 44/60 of the three-dimensional electron-density map of hen egg-white lysozyme at 2 Å resolution. AA′ shows the axis of a length of α-helix lying in the plane of the sections. B indicates an α-helix more nearly normal to the sections. C indicates the disulfide bridge between residues 30 and 115, the sulfur atoms of which lie one above the other. The side chain of a phenylalanine residue is located four residues along the helix from the disulfide, towards the lower sections. Reproduced with permission from *Nature* (Blake *et al.*, 1965). Copyright (1965) Macmillan Magazines Limited.

26.1.3.10. *The electron-density map at 2 Å resolution*

GAM joined the team in early 1964 with the specific task of writing a program for calculating the 2 Å electron-density map. Fortunately, at this stage, the University of London Computing Centre was in the course of acquiring a Ferranti ATLAS computer, which was then one of the most powerful computers available for this kind of work, and GAM eagerly set about taking advantage of its power. Whereas we ourselves had operated the previous MERCURY computer at night, there were machine operators for the ATLAS with whom we left our input tapes; these were very large reels of paper tape comprising all of the amplitude and phase data for the 2 Å set. Following two or three unsuccessful attempts to load the whole of the data without tearing the tapes, GAM modified his program so that the data could be loaded in several sections, which was achieved satisfactorily.

The electron density was calculated at 1/120ths of the cell edge along *a* and *b*, and 1/60ths along *c*. The output from the computer was on punched paper tape, arranged with appropriate insertions of carriage returns and line feeds so that the teleprinter output was in a form suitable for immediate contouring to a scale of 0.75 inches equal to 1 Å. Each *x*, *y* section of the map was printed out in five strips, which had to be glued together to cover the whole area. The contours were drawn initially in pencil on these paper sheets and were then copied in drawing ink to thin Mylar sheets, which were supported for interpretation on Perspex sheets that were stacked and bolted together, with spacers of appropriate dimensions to maintain the scale in the *c* direction, in groups of five for ease of handling. The whole map was drawn on 60 sections perpendicular to the *z* axis, and the bolts holding the blocks of five sheets together were designed to fit into one another to keep successive blocks in register. The maps were viewed on large light boxes, specially constructed for the purpose, though they were not transparent enough for more than fifteen sheets, three blocks of five, to be studied in detail at one time. The contours were drawn on Mylar sheets to avoid waste of the more expensive Perspex that would have arisen from errors. A grid was drawn on a Mylar sheet, which could be superimposed on the contour stacks and used to read the atomic coordinates directly in ångstroms.

The electron density had been calculated using a scale factor such that it was convenient to draw contours at intervals of 0.25 e Å^{-3}; as no $F(000)$ term had been included in the calculation, electron-density values were relative to the mean value for the unit cell. Contours were drawn only for electron densities above the mean, the two lowest levels being drawn in orange ink, and the higher ones in black ink. This proved to be a satisfactory form of representation, as it very clearly revealed regions of continuous high electron density while also showing the positions of significant features of lower density. All the members of the lysozyme group participated in drawing the maps.

To illustrate the result that was obtained, sections $z = 35/60$ to 44/60 of the electron-density map are shown in Fig. 26.1.3.9.

26.1.3.11. *Map interpretation and model building*

We were fortunate that by the time the map was ready for interpretation, two independent groups of protein chemists, led by Pierre Jollès in Paris and R. E. Canfield in New York, had studied the amino-acid sequence of hen egg-white lysozyme in detail and published their results. Two slightly different amino-acid sequences, complete with the arrangement of the four disulfide bonds, were published by Jollès *et al.* (1964) and by Canfield & Liu (1965), and this information was used intensively in the interpretation of the electron-density map of the protein. The sequence published by Canfield & Liu (1965) is shown in Fig. 26.1.3.10.

Many features were immediately identifiable in the map. These included the side chains of many of the amino-acid residues, especially the disulfide bridges and the aromatic side chains of tryptophan, tyrosine and phenylalanine. Nevertheless, no attempt was made to interpret the map in detail without recourse to the amino-acid sequence. Interpretation began in the part of the map

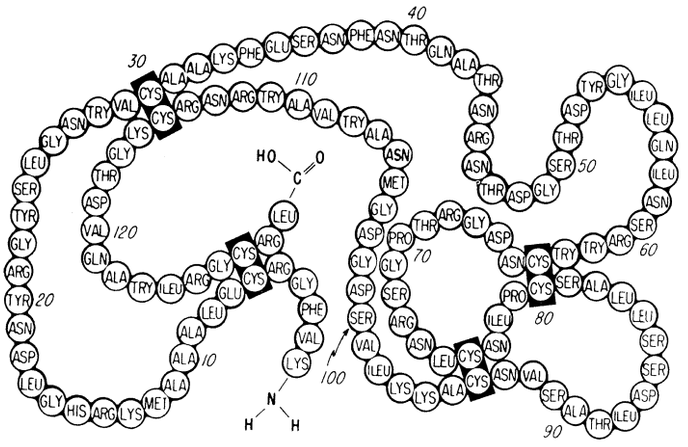

Fig. 26.1.3.10. The amino-acid sequence of hen egg-white lysozyme (Canfield & Liu, 1965).

shown in Fig. 26.1.3.9. The strongest feature in this part of the map corresponded to a disulfide bridge (C), in which the two sulfur atoms lie one above the other in the direction of the *c* axis. The first challenge was to identify this bridge. It is clearly connected to a helical region of the molecule (A), which ran in the direction from top right to bottom left of the diagram, with the main-chain carbonyl groups pointing in this direction. Consequently, this helix ran from its amino terminus on the right to its carboxyl terminus in the centre of the map. The map was quite clear enough to count the α-carbons from the cysteine residue that forms part of the disulfide bridge, and it was immediately apparent that the fourth residue from the cysteine towards the carboxyl terminus is an aromatic residue, probably phenylalanine. Inspection of the amino-acid sequence in Fig. 26.1.3.10 showed that only one pair of residues satisfied this condition, Cys30 and Phe34.

Given this start, interpretation of the map and the construction of a molecular model were relatively straightforward. The model was constructed in a metal frame, the top and bottom of which consisted of sheets of blockboard. The *a* and *b* axes were drawn parallel to the diagonals of these boards to cover the coordinate ranges, respectively, $x = -1/4$ to $+1/4$ and $y = 0$ to $+1/2$ to a scale of 2 cm to 1 Å. This was the scale of the brass models, constructed by Cambridge Repetition Engineers Ltd, which were used to build the model. The height of the frame covered the full extent of the *c* axis. The heavy-atom coordinates and the computer programs were both based on the wrong-handed space group $P4_12_12$. It was not until the anomalous scattering from the heavy atoms was incorporated that the correct space group $P4_32_12$ was assigned. The Fourier-map sheets were actually stacked the opposite way round and a left-handed system of axes was used for the model. In retrospect, this should have been put right at once, but the system was not easy to change.

Holes were drilled in the top and the base boards on the grid defined by the *a* and *b* axes, and these were used to support an array of brass rods parallel to the *c* axis to which the model components could be attached. The model building was carried out by two subgroups, CCFB and VRS in one and ACTN and DCP in the other, so that work could go on continuously throughout each day. The method employed was to examine the map density corresponding to the next amino-acid residue to be located and to mark the positions of the constituent atoms with small washers or nuts. The coordinates of these atoms were then read from the map (making use of the superimposed grid and estimating the *z* coordinates from the extent to which adjacent *z* sections contributed to the density). These

Table 26.1.3.3. *Discrepancies in amino-acid sequences (excluding Asp/Asn)*

	Residue						
Reference	40	41	42	58	59	92	93
Canfield & Liu (1965)	Thr	Gln	Ala	Ile	Asn	Val	Asn
Jollès *et al.* (1964)	Gln	Ala	Thr	Asn	Ile	Asn	Val

coordinates were then located in the model by means of the coordinate grids drawn on the base and top boards and by the use of a plumb line marked with the *z* coordinates. At this stage it was usually possible to fix a model of the amino-acid residue in place in the model frame with remarkably little trouble, though, of course, fine adjustment was necessary as the model grew.

This bout of model building began towards the end of February 1965 and proceeded quite rapidly. The main difficulties arose from the fact that the two amino-acid sequences that were available did not agree in every respect. Eleven amino-acid residues were identified differently by Jollès (Jollès *et al.*, 1964) and by Canfield (Canfield & Liu, 1965). Four of the discrepancies involved Asn and Asp, which cannot be distinguished in the electron-density map with any degree of certainty. The remaining discrepancies are shown in Table 26.1.3.3.

Inspection of the residues 40, 41, 42, 92 and 93 showed quite clearly that the shapes in the electron-density map fitted the Canfield side chains. Our initial conclusion, however, was that the electron densities corresponding to residues 58 and 59 were more consistent with the sequence proposed by Jollès than that published by Canfield. Accordingly, in our first detailed description of the

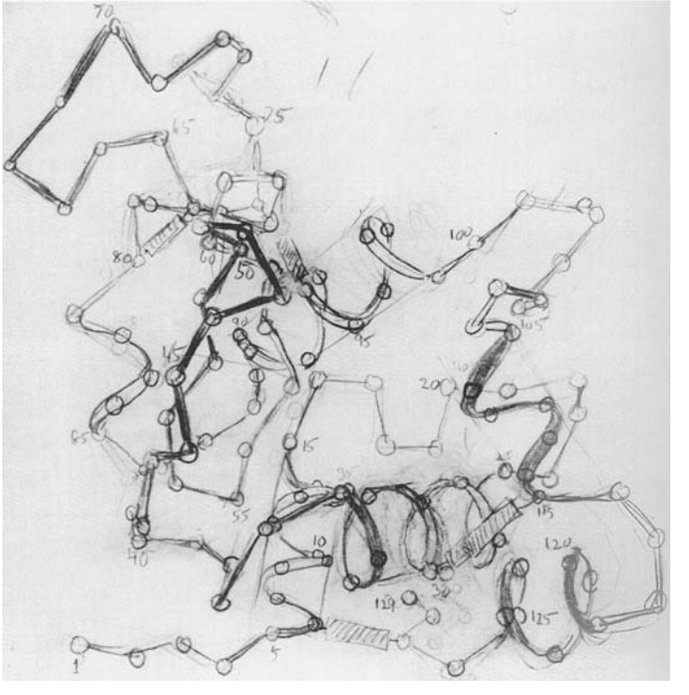

Fig. 26.1.3.11. Schematic drawing of the main-chain conformation of lysozyme. The drawing was made from observations of the molecular model by Sir Lawrence Bragg and later prepared for publication by Mrs S. J. Cole.

structure (Blake *et al.*, 1965), we accepted the identification of these residues by Jollès and his colleagues. As will be seen, this was a mistake.

The model was completed in time for it to form the centrepiece at Bragg's 75th birthday party at the Royal Institution on 31 March 1965, and a description was submitted to *Nature* at about the same time (Blake *et al.*, 1965). One of the difficulties at this stage of development of the subject was that computer graphics had not yet been developed to the stage at which illustrations of protein structures could be produced with any degree of facility. Consequently, we were most grateful that Sir Lawrence Bragg enthusiastically drew the model freehand. His drawing provided the main illustration for the published paper: the original sketch is reproduced in Fig. 26.1.3.11.

During the early summer of 1965, we rebuilt the model, taking care to ensure that the components fitted the electron density as well as could be judged by eye, and that the contacts between them were fully consistent with current understanding of van der Waals forces and hydrogen bonds. The model components were fixed firmly in position with small brass clamps, designed and constructed by Bruce Morris in the Royal Institution workshop, so that at the end of the process we had an unrealistically rigid model that incorporated all the information available to us at the time. At this stage it became clear that the Canfield sequence (Fig. 26.1.3.10) was to be preferred in every respect.

A stereo-photograph of this model (Blake *et al.*, 1967) is shown in Fig. 26.1.3.12 in a view that shows the most striking feature of the molecule. The molecular face in the foreground of this view is crossed by a deep cleft, roughly parallel to the *c* axis of the crystal unit cell.

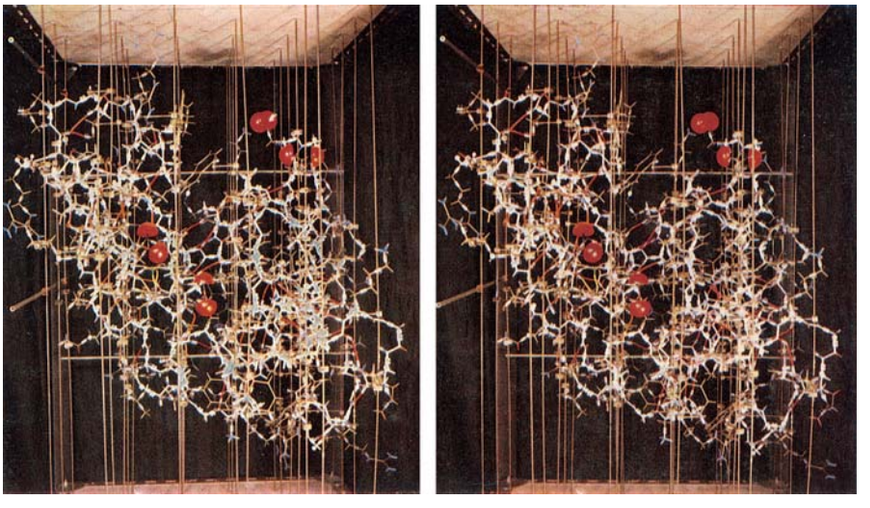

Fig. 26.1.3.12. Stereo-photographs of a model of the lysozyme molecule to a scale of 2 cm to 1 Å. The main polypeptide chain is painted white, and nitrogen, oxygen and sulfur atoms are indicated by blue, red and green sleeving, respectively. Some hydrogen bonds are shown by red connections. Oxygen atoms of the acid side chains near the cleft, Glu 35, Asp 52, Asp 101 and Asp103, are shown by red hemispheres (Blake *et al.*, 1967).

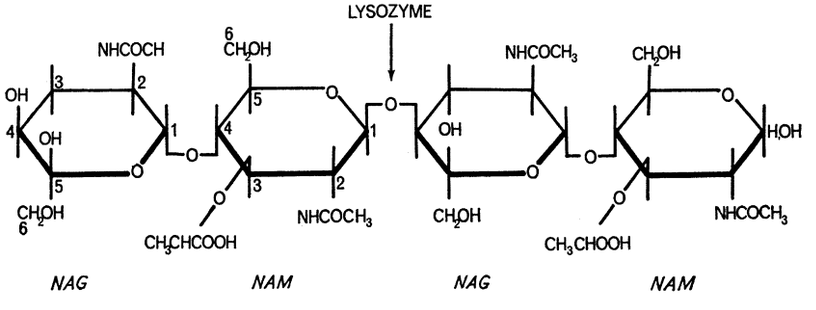

Fig. 26.1.4.1. The cell-wall tetrasaccharide with the $\beta(1\rightarrow4)$ glycosidic bond that is hydrolysed by lysozyme indicated (Blake *et al.*, 1967).

26.1.4. Structural studies on the biological function of lysozyme

26.1.4.1. *Lysozyme substrates*

By the summer of 1964, the results for the structure determination of lysozyme were sufficiently promising for us to consider diffraction studies on the biological function of lysozyme. At about the same time, the Royal Society invited Max Perutz to organise a Discussion Meeting on lysozyme at The Royal Institution in early 1966, to which the leading workers worldwide would be invited. This provided an additional spur to our efforts.

Simultaneously with his discovery of lysozyme in 1922, Fleming had found a gram-positive species of bacteria, *Micrococcus lysodeikticus*, that was particularly susceptible to lysis by lysozyme (Fleming, 1922). During the 1940s and 1950s, further work by a number of authors had shown that lysozyme exerted its biological effects through hydrolysis of the bacterial cell wall. These studies led to a definitive description of the polysaccharide component of the bacterial cell wall, as shown in Fig. 26.1.4.1.

The structure is composed of alternating sugars, linked by $\beta(1\rightarrow4)$ glycosidic links, of *N*-acetylglucosamine (GlcNAc, abbreviated by us at the time as NAG) and *N*-acetylmuramic acid (MurNAc, previously abbreviated as NAM). The 3-hydroxyls of the MurNAc residues are attached to short peptides, and the peptides themselves are cross-linked to provide an extensive and rigid two-dimensional proteoglycan network. Lysozyme cleaves the $\beta(1\rightarrow4)$ glycosidic bonds between MurNAc and GlcNAc residues, thus leading to the dissolution of the bacterial coat and lysis of the bacterium.

Kinetic studies on the activity of lysozyme were hampered by the lack of a suitable small-molecular-weight substrate. The turbidometric assay in use at the time, and still in use today, followed the change in optical density of a suspension of *Micrococcus lysodeikticus* cells as the cells were lysed by lysozyme. The assay could work reliably but it was sensitive to physical parameters such as ionic strength, the method by which the cells were suspended and product inhibition. In 1962, Wenzel *et al.*, in an effort to obtain a small-molecular-weight substrate, had reported that lysozyme promoted the cleavage of the trimer of GlcNAc, tri-*N*-acetylchitotriose, releasing dimer and monomer sugars, and that the monomer, GlcNAc, was an inhibitor of lysozyme (Wenzel *et al.*, 1962). Compounds such as glucose or cellobiose that lacked the *N*-acetyl group did not inhibit.

John Rupley extended this work. Rupley, a chemist from the University of

references