KURIHARA, T., UCHIDA, A., OHASHI, Y., SASADA, Y. & OHGO, Y. (1984). *J. Am. Chem. Soc.* **106**, 5718–5724.
MIRSKY, K. (1978). *Computing in Crystallography, Proceedings of an International Summer School in Crystallographic Computing*, p. 169. Delft Univ. Press.
OHASHI, Y. (1988). *Acc. Chem. Res.* **21**, 268–274.
OHASHI, Y., SASADA, Y., TAKEUCHI, S. & OHGO, Y. (1980). *Bull. Chem. Soc. Jpn*, **53**, 1501–1509.
OHASHI, Y., TOMOTAKE, Y., UCHIDA, A. & SASADA, Y. (1986). *J. Am. Chem. Soc.* **108**, 1196–1202.
OHASHI, Y., UCHIDA, A., SASADA, Y. & OHGO, Y. (1983). *Acta Cryst.* B**39**, 54–61.
OHASHI, Y., YANAGI, K., KURIHARA, T., SASADA, Y. & OHGO, Y. (1981). *J. Am. Chem. Soc.* **103**, 5805–5812.
OHASHI, Y., YANAGI, K., KURIHARA, T., SASADA, Y. & OHGO, Y. (1982). *J. Am. Chem. Soc.* **104**, 6353–6359.
TOMOTAKE, Y., UCHIDA, A., OHASHI, Y., SASADA, Y., OHGO, Y. & BABA, S. (1985). *Isr. J. Chem.* **25**, 327–333.
UCHIDA, A., OHASHI, Y., SASADA, Y., OHGO, Y. & BABA, S. (1984). *Acta Cryst.* B**40**, 473–748.

Acta Cryst. (1990). B**46**, 54–62

Refinement of Triclinic Lysozyme: I. Fourier and Least-Squares Methods

BY JOHN M. HODSDON, GEORGE M. BROWN,* LARRY C. SIEKER AND LYLE H. JENSEN

Departments of Biological Structure and Biochemistry, University of Washington, Seattle, Washington 98195, *USA*

(*Received* 15 *February* 1989; *accepted* 7 *August* 1989)

Abstract

X-ray diffraction data to 1·5 Å resolution have been collected for triclinic crystals of hen egg white lysozyme. The triclinic model was derived from the tetragonal one by the rotation function and refined initially by $F_o - F_c$ and differential difference syntheses against 2 Å resolution data. Refinement was continued by differential difference cycles against the 1·5 Å data until R was reduced to 0·220. Although the initial refinement was rapid, it was subsequently a matter of attrition, leading to a complete recheck of the data and the discovery of systematic error which affected primarily the high-resolution data. Refinement was continued against the corrected 2 Å data by block-diagonal least squares. After five cycles the refinement was terminated at $R = 0{\cdot}254$ because of the imminent availability of a preferred refinement program. Problems with the protein model, the solvent, and the interaction of the scale and thermal parameters are discussed. The experiences gained in this study are summarized.

Introduction

Lysozyme is an enzyme, widely distributed in biological systems, which catalyzes the hydrolysis of polysaccharides in the bacterial cell wall. The particular lysozyme found in the whites of hen eggs has a single polypeptide chain of 129 amino acids, cross-linked by four disulfide bridges. The molecule has 1001 nonhydrogen atoms and a mass of 14 300 daltons. Hen egg white (HEW) lysozyme crystallizes from aqueous solutions in at least four different crystal systems (Steinrauf, 1959), and the structure of the tetragonal form has been determined by the multiple-isomorphous-replacement method (Blake, Koenig, Mair, North, Phillips & Sarma, 1965).

Triclinic crystals of HEW lysozyme are unusual in that they have a relatively low solvent content, 26% by weight. They are also unusual in the sense that they diffract to high resolution, considerable intensity being observed for reflections to d spacings of at least 1·0 Å. It is clear, therefore, that sufficient data can be observed to yield a precise model of the molecule in this crystal form. Such a model will be useful in determining the extent of departure from ideal bond and torsion angles in proteins, in studying the solvent structure in protein crystals, in attempting to visualize the effects of radiation damage to triclinic lysozyme, in comparative studies of the effects of packing forces on the molecular structure of lysozyme in different crystal forms, and as the initial model for neutron diffraction studies of triclinic lysozyme. When this work was initiated (1971), it was not clear to what extent protein models could be refined nor the most effective way of doing so. One of our objectives, therefore, was to determine what was feasible in a favorable case and what could be learned about the methods of refinement and their limitations.

This paper provides an account of the initial efforts to derive an acceptable model for triclinic lysozyme.

* Permanent address: Division of Chemistry, Oak Ridge National Laboratory, Oak Ridge, Tennessee, USA.

0108-7681/90/010054-09$03.00

Experimental

Triclinic crystals of HEW lysozyme were grown from 1% aqueous lysozyme solution containing 2% $NaNO_3$, buffered with sodium acetate at pH 4·55 (Kurachi, Sieker & Jensen, 1976). The unit-cell parameters are the following: $a = 27{\cdot}283$ (10), $b = 31{\cdot}980$ (10), $c = 34{\cdot}291$ (13) Å, $\alpha = 88{\cdot}53$ (5), $\beta = 108{\cdot}57$ (3), $\gamma = 111{\cdot}85$ (3)°, based on λ(Cu $K\alpha$) = 1·5418 Å. These values are averages from eight different crystals, the estimated standard deviations being calculated from the population variance. The space group is $P1$, with one molecule in the unit cell.

Intensity data were collected on a computer-controlled, four-circle diffractometer, operating in the $\omega/2\theta$ scan mode. The Cu-target X-ray tube with focal spot 0·4 × 10 mm, operating at 40 kV and 30 mA, was set at a 3° take-off angle and the radiation filtered through 0·009 mm Ni foil in the incident beam. The pulse-height analyzer passed approximately 95% of the Cu $K\alpha$ photons.

Using four crystals, we collected the full sphere of data to a d spacing of 1·54 Å. Table 1 lists the relevant crystal data. At a room temperature of approximately 295 K, crystals deteriorate 25–30% in the X-ray beam in approximately four days, as determined from 13 standard reflections measured at intervals of ~500 reflections. For each crystal, reflections were collected in a given order for one hemisphere, then in the reverse order for the second hemisphere. Intensities were corrected for coincidence loss, for absorption (North, Phillips & Mathews, 1968) and for crystal deterioration.

Table 1. *Crystal data*

Crystal No.	Range of d (Å)	No. of reflections	D_F* Friedel pairs
1	∞–2·50	7838	0·012
2	2·56–1·97	7962	0·021
3	1·99–1·70	8807	0·026
4	1·71–1·54	8299	0·037

Crystals	D_F common reflections	No. of pairs
1/2	0·021	250
2/3	0·014	213
3/4	0·020	221

*$D_F = \Sigma||F_i| - |\bar{F}||/\Sigma|F_i|$.

Refinement

Initial model refinement

The beginning coordinates were those from the tetragonal structure which had been idealized by Diamond's model-building program [designated *RS5D* by Diamond (1974)]. They were transformed to the triclinic cell by taking $\chi = 49°$, $\varphi = 98°$, $\psi = 85°$ as given by Joynson, North, Sarma, Dickerson & Steinrauff (1970), except for the sign of χ which was taken as negative. Use of the positive sign given in the paper led to erroneous structure factors.

In checking the triclinic model derived in this way for close intermolecular approaches, we found six side chains with atoms separated by distances less than 1·5 Å, four of the six residues being arginine. Such impossibly close contacts are not suprising, however, since the model was derived from a different crystal form, but they do show that the initial model had substantial error.

The first structure-factor calculation was based on the 1001 atoms in the lysozyme molecule with an overall B of 5 Å^2. The conventional R (= $\Sigma||F_o| - |F_c||/\Sigma|F_o|$) was 0·508 for the 7142 reflections with $10 > d > 1{\cdot}97$ Å (those with $d > 10$ Å were weighted zero). In order to monitor R as a function of sin θ/λ, the data set was divided into five groups with d spacings in the ranges ∞–10, 10–5, 5–3, 3–2·5 and 2·5–1·97 Å. The uppermost plot in Fig. 1 shows R as a function of sin θ/λ for the first structure-factor calculation. The point for data in the range $0{\cdot}0 < \sin\theta/\lambda < 0{\cdot}05$ Å^{-1} is not shown because we expected the omitted solvent to have a very large effect on the low-angle reflections (Watenpaugh, Sieker, Herriott & Jensen, 1972).

The broken curve in Fig. 1 is based on Luzzati's theoretical relation for a mean error of 0·9 Å (Luzzati, 1952). Although the upper plot for the initial structure-factor calculation does not follow the theoretical curve well, it suggests that the model had relatively large errors.

The initial refinement cycles were restricted to the 2 Å resolution data set, and the first two cycles were by $F_o - F_c$ syntheses. In the first one, R decreased to

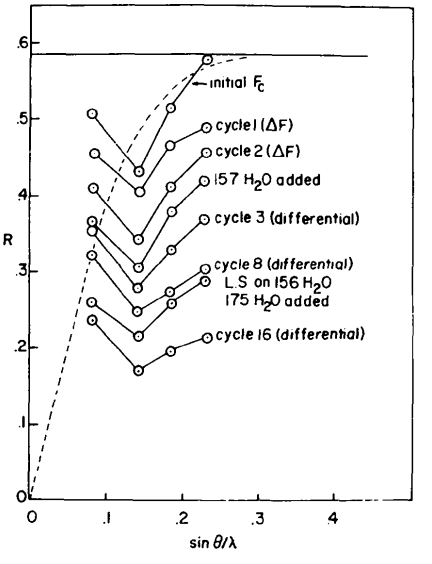

Fig. 1. R *vs* sinθ/λ (Å^{-1}) at the indicated stage in the initial refinement against 2 Å resolution data.

0·452 after adjusting the overall B to 10 Å^2 and idealizing the model (Hermans & McQueen, 1974); in the second cycle, R decreased to 0·403 without model idealization (see Figs. 1 and 2). In both $F_o - F_c$ maps, many prominent peaks were evident in the solvent spaces. Accordingly, 157 fully occupied oxygen atoms with $B = 15$ Å^2 were added to the model, representing approximately half the water in the crystal and reducing R = to 0·368.

In order to expedite the refinement, we continued by an essentially automatic method, differential difference syntheses (Booth, 1946; Cochran, 1951). The model was idealized periodically to maintain appropriate bond lengths and angles (Freer, Alden, Carter & Kraut, 1975; Hermans & McQueen, 1974), calculating $F_o - F_c$ and $2F_o - F_c$ maps only as required to check the model for evidence of errors. R decreased rapidly at first, reaching 0·289 after two differential difference cycles (four refinement cycles, Fig. 2), but thereafter much more slowly.

After the eighth refinement cycle, occupanices and B values of 156 solvent oxygen atoms were adjusted by least squares (one of the 157 orginally included in the model was inadvertently omitted). R decreased only from 0·281 to 0·276. At that point adding 175 water oxygen atoms with partial occupancies ranging from 0·5–1·0 and B values of 15 Å^2 reduced R to 0·256.

As noted above, the coordinates from the first refinement cycle were idealized, but thereafter coordinates were idealized only after every third cycle. Typically, R increased by 0·06 to 0·07 on idealization, and in the early stages the relatively large increases were attributed to the rather rigid restraints applied in idealizing the model. These were relaxed somewhat at the 16th cycle, but R still increased from 0·198 to 0·256 on idealization at that point. After the final differential difference cycle based on the 2 Å data set (cycle 17), R was 0·201 without idealization.

Although R had decreased to a value generally considered acceptable at the time (1973), it was still much higher than the apparent precision of the data, and the refinement had essentially converged. Despite the apparently acceptable value of R, the behavior of the refinement, particularly the relatively large increase in R on idealization, suggested that substantial error still remained in the model, but it was not clear either in $F_o - F_c$ or in $2F_o - F_c$ maps how to correct difficult regions. We decided, therefore to continue the refinement with a more extensive data set, adding the 7781 reflections to 1·54 Å resolu-

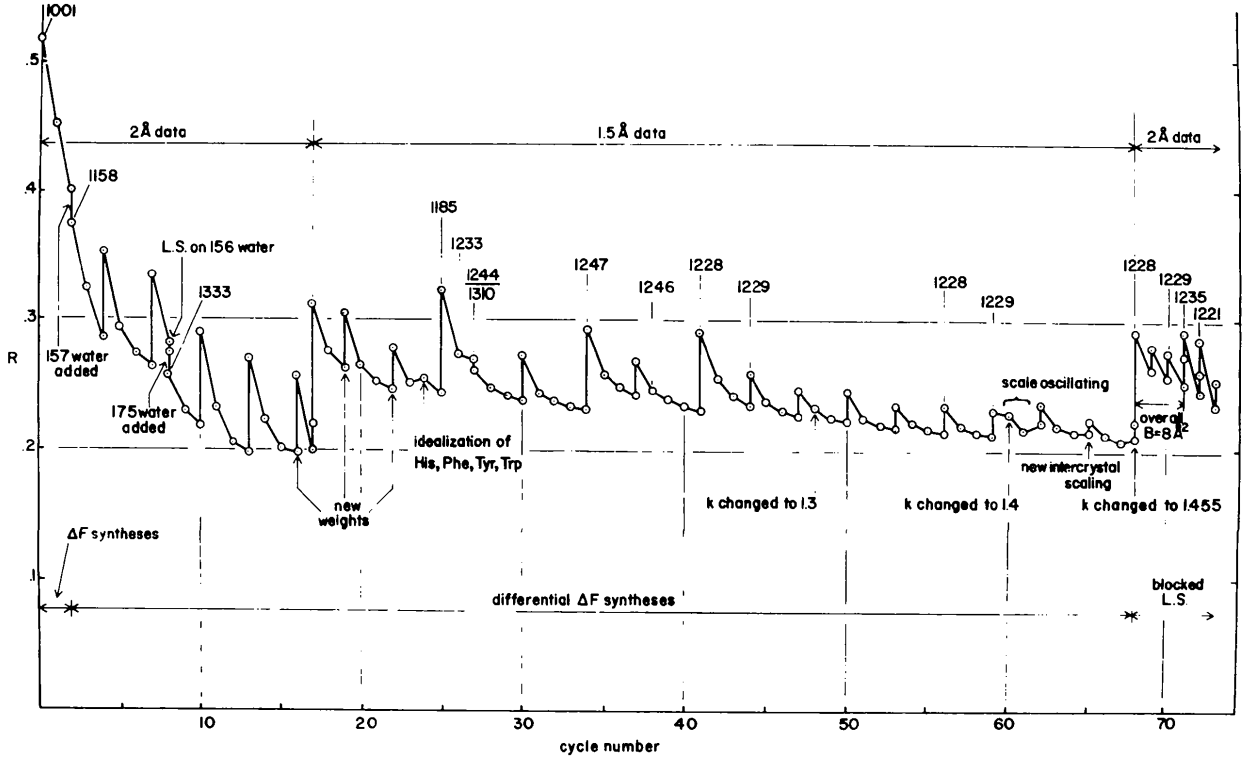

Fig. 2. R *vs* cycle number. Unless otherwise indicated, increase in R at a given cycle results from idealizing the model, sometimes combined with editing. Each number above the plot and joined to a given point, *e.g.*, 1158 at cycle 2, is the total number of atoms in the model at that point, and it remains the same until a different number appears.

tion. The data beyond 1·97 Å were divided into two groups with d spacing in the ranges 1·97–1·71 and 1·71–1·54 Å, the set of data with $10 > d > 1{\cdot}54$ Å comprising 14 923 reflections. This set will be referred to as the 1·5 Å data set.

Refinement continued through a series of 51 differential difference cycles (Hodsdon, Sieker & Jensen, 1975). In Fig. 2 it is evident that the refinement against the 1·5 Å data was a matter of attrition: the 51 cycles reduced R only from 0·312 to 0·220 (idealized). Although R did not decrease rapidly, the model changed substantially. Much of the checking, particularly of external side chains and solvent that characterized this part of the investigation, does not warrant detailed description, and we include here only the features which serve to transmit the essential behavior of the refinement.

The coordinates from cycle 17 were used in the first structure-factor calculation for the 1·5 Å data set. The plot of R *versus* $\sin\theta/\lambda$ shows a pronounced discontinuity between the last group of reflections in the 2 Å data set and the two groups of added data (see Fig. 3, plot for cycle 17). In fact, R for the latter groups approaches the value of 0·59 expected for a random distribution of atoms. This was a disquieting feature of the refinement to this point, but it was not the only observation that caused concern. We also found that after scaling the data following each refinement cycle so that $k\Sigma F_o = \Sigma F_c$ for the whole data set, the rescale factors for each group of reflections, $k' = \Sigma F_c(\text{group})/\Sigma F_o(\text{group})$, varied considerably and apparently systematically. Fig. 4 shows plots of k' as a function of $\sin\theta/\lambda$ after cycle 17, the end of the initial refinement with 2 Å data, and after the first cycles with the 1·5 Å data set, cycle 21.

In the plot for cycle 17 in Fig. 4, the trend in k' beyond $\sin\theta/\lambda = 0{\cdot}1$ Å^{-1} shows that ΣF_c decreases less rapidly than ΣF_o with increasing $\sin\theta/\lambda$, a trend that can be compensated by increasing the temperature factor. Accordingly, the overall scale factor k for scaling F_o on input to each refinement cycle was allowed to 'float', and the B values were adjusted as dictated by the data with $\sin\theta/\lambda > 0{\cdot}1$ Å^{-1}, reflections with $\sin\theta/\lambda < 0{\cdot}1$ Å^{-1} being weighted zero. In the four cycles from 18 to 21, k decreased from 1·50 to 1·25 and the systematic trend in k' beyond $\sin\theta/\lambda = 0{\cdot}1$ Å^{-1} was eliminated with a compensating increase in mean B from ~11 to ~14 Å^2. As expected, however, k' for the group of reflections with $\sin\theta/\lambda < 0{\cdot}1$ Å^{-1} was now much greater than unity (Fig. 4, plot for cycle 21).

Since $d_{\min}$ for the 1·5 Å data set is near the distances separating convalently bonded C, N and O atoms, the restraints in the idealization program were further relaxed by reducing the weights by a factor to two in cycle 19 and another factor of two in cycle 22, causing the r.m.s. deviations from ideal values to fall in the range 0·04–0·05 Å for bond lengths and 5–8° for bond angles.

On inspecting both $F_o - F_c$ and $2F_o - F_c$ maps after cycle 24, it was clear that the model suffered errors greater than we had anticipated on the basis of the 2 Å refinement. Not only did the electron density or difference density in a number of side chains, particularly those of arginine, indicate model errors had not yet been corrected, but the electron density for the main chain from Asn65 through Arg73 was relatively low and in some regions diffuse. In fact, Pro70 could not be recognized in either $F_o - F_c$ or $2F_o - F_c$ maps, and difficulty was evident at Gly71 and Ser72 as well, but how to correct the model was not evident. Beginning at cycle 25, therefore, extensive

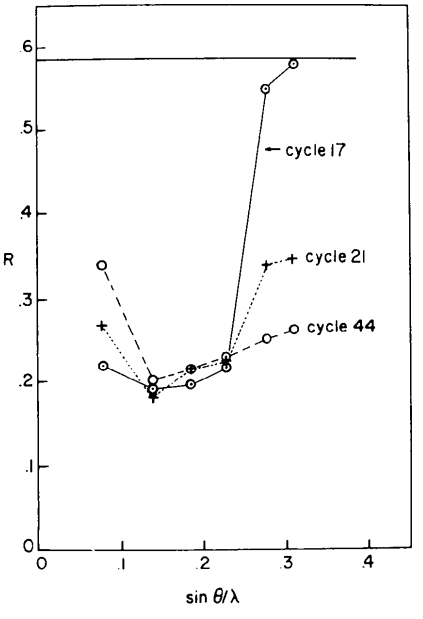

Fig. 3. R *vs* $\sin\theta/\lambda$ (Å^{-1}) for cycles 17 and 21 and 44.

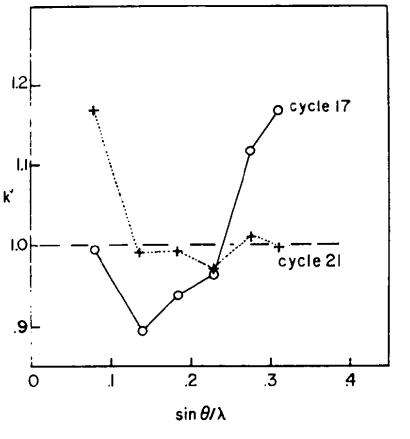

Fig. 4. k' *vs* $\sin\theta/\lambda$ (Å^{-1}) for cycles 17 and 21 (see text).

editing was initiated by deleting many of the solvent atoms, and at intervals through the next 34 cycles by adjusting atoms or groups of atoms in the protein and adjacent solvent on the basis of $F_o - F_c$ and $2F_o - F_c$ maps, while at the same time checking contact distances. Occasionally, groups of atoms were removed for one or more cycles to relax bias in the phases from incorrect positions. The numbers above the plot and attached to particular cycles in Fig. 2 refer to the total number of atoms in the model at the indicated points.

When the overall scale factor k was allowed to float, it decreased to 1·163, and the mean B increased to ~16 Å^2, a value which, although it fitted the data beyond $\sin\theta/\lambda = 0{\cdot}1$ Å^{-1}, was clearly too large when the falloff of the intensities was compared, for example, with rubredoxin from *C. pasteurianum* which has an overall $B = 12$ Å^2 (Watenpaugh, Sieker, Herriot & Jensen, 1972). Accordingly, k was increased to 1·3 at cycle 48 and to 1·4 at cycle 60. Plots of k' *versus* $\sin\theta/\lambda$ at these two cycles are shown in Fig. 5. Both plots show a distribution of k' similar to that for cycle 17 in Fig. 4.

Data check

The tedious pace of the refinement and the question concerning the behavior of k' led to a reevaluation of the data, involving a redetermination of the intercrystal scaling constants for the four crystals used for the 1·5 Å data set (see Fig. 2, cycle 65) and a careful recheck of every aspect of the data-reduction process.

Data from the four crystals which constitute the 1·5 Å set had been scaled together by common reflections in the overlap regions between the successive shells. Since this could have led to a sytematic trend in scaling the data, a different method was used in the redetermination. Two zones of data (the *hk*0 and *h*0*l* which cut across all shells of the data set) were collected from a crystal, and the reflections common to each of the shells were used for scaling. The new factor scaling crystal 1 to 2 differed from the original one by 6%, but the factors for crystals 2 to 3 and 3 to 4 differed by less than 1% from the original ones. The changes in the scale factors were sufficient to account, only in part, for the effects seen in Fig. 4 for cycle 17 and in Fig. 5.

Data for the two zones of the crystal used for redetermining the intercrystal scale factors were collected to d spacings of 1·01 Å and approximately 1000 reflections in each zone were used to determine B values (Wilson, 1942). The values were 7·8 Å^2 for the *hk*0 zone and 8·4 Å^2 for the *h*0*l* zone, somewhat less than the value of 10 Å^2 assumed after the first refinement cycle.

In rechecking the data, the ratio $\sum I_{\overline{hkl}}/\sum I_{hkl}$ was plotted for each crystal in groups of 100 pairs of reflections as a function of the order in which the reflections were collected in the first hemisphere. The ratio should be near unity. This was found to be true for crystal 1 as shown in Fig. 6(*a*), but for crystals 2–4 (data with $d < 2{\cdot}5$ Å) systematic deviations were found which increased with successive crystals, *i.e.*,

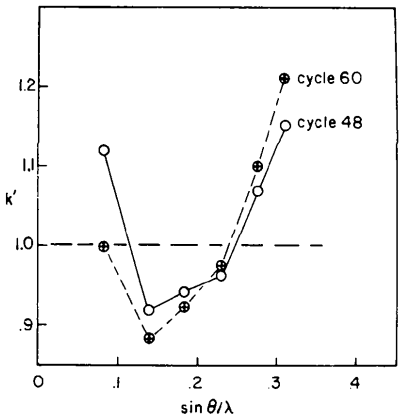

Fig. 5. k' *vs* $\sin\theta/\lambda$ (Å^{-1}) for cycles 48 and 60 (see text).

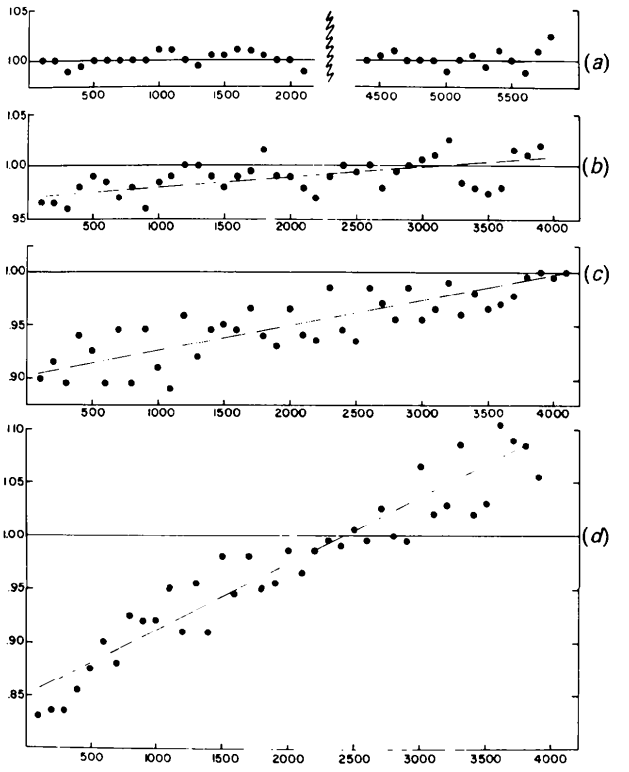

Fig. 6. $\sum I_{\overline{hkl}}/\sum I_{hkl}$ for groups of 100 reflections *vs* serial number of reflections in first hemisphere. (*a*) Crystal 1, data collected in two shells, ∞–3 Å followed by 3–2·5 Å. (*b*) Crystal 2, 2·56–1·97 Å. (*c*) Crystal 3, 1·99–1·70 Å. (*d*) Crystal 4, 1·71–1·54 Å.

with increasing $\sin\theta/\lambda$, Figs 6(*b*)–6(*d*). The trend in the deviations for crystal 2, Fig. 6(*b*), was considered to be sufficiently small that we were justified in scaling the intensities in the second hemisphere to match those in the first by use of a linear function derived from the plot. The trend for crystals 3 and 4 shown in Figs. 6(*c*) and 6(*d*) is much more serious, and for lack of a definitive explanation, we did not feel justified in continuing to use the data from these crystals. Accordingly, we terminated the refinement against the 1·5 Å data set after cycle 68.

Since we intended to continue refining against the corrected 2 Å data by least squares, it was essential that we have appropriate weights. In calculating standard deviations in the intensities, we used the expression $\sigma(I) = [\sigma_c^2(I) + (KI)^2]^{1/2}$, where $\sigma_c(I)$ is the error from counting statistics and K is a constant to be determined for a given crystal and experimental facility (Busing & Levy, 1957). Suitable values of K for crystals 1 and 2 were determined from normal probability plots (Abrahams & Keve, 1971) by comparing for each crystal the expected values of Δ_{hkl} against the experimental values for several different K, where $\Delta_{hkl} = (I_{hkl} - I_{\overline{hkl}})/\sigma^2_{hkl} + \sigma^2_{\overline{hkl}})^{1/2}$. Figs 7(*a*) and 7(*b*) are plots for crystal 1 with $K = 0{\cdot}04$ and 0·05. The former closely represents the scatter in the data. In a similar way K was determined to be 0·07 for crystal 2. Figs 8(*a*) and 8(*b*) are normal probability plots for crystal 2 before and after correcting for the trend shown in Fig. 6(*b*).

For the rescaled, corrected 2 Å data from crystals 1 and 2, the intensities of 7075 reflections in the range 10–1·97 Å exceeded $2\sigma(I)$ compared with 7142 reflections in the earlier set.

Least-squares refinement against the corrected 2 Å data set

Five least-squares refinement cycles against the corrected 2 Å data were calculated, numbered 69–73 in Fig. 2, idealizing the model after each cycle. A value of 1·455 was determined for the scale constant, and an overall $B = 8$ Å^2 was used in cycles 69–71 before applying individual B values for each atom in the last two cycles. In addition to idealizing the protein after cycles 71 and 72, the water model was

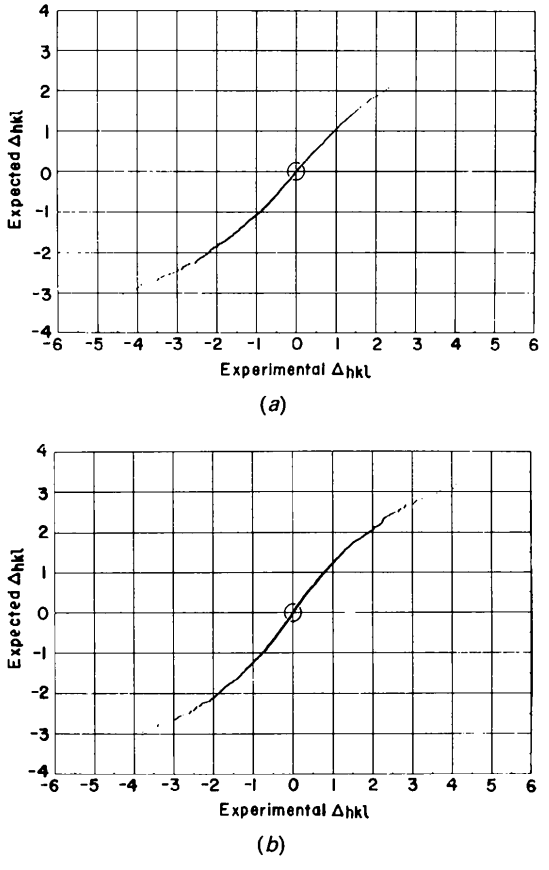

Fig. 7. Normal probability plot of the expected value of $(I_{hkl} - I_{\overline{hkl}})/(\sigma^2_{hkl} + \sigma^2_{\overline{hkl}})^{1/2}$ *vs* the experimental value for (*a*) $K = 0{\cdot}04$, (*b*) $K = 0{\cdot}05$.

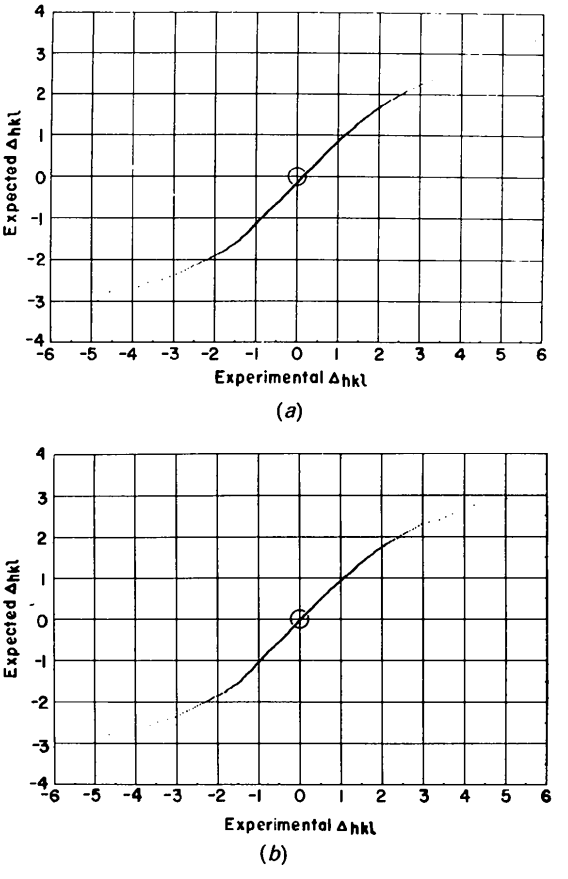

Fig. 8. Normal probability plot as in Fig. 7 for data from crystal 2. (*a*) Before data corrected for systematic trend. (*b*) After data corrected, both plots for $K = 0{\cdot}07$.

extensively checked with a net increase of six water oxygen atoms after cycle 71 and a net decrease of 14 after cycle 72.

During these least-squares cycles, R decreased steadily, but it was still at a relatively high value, 0·254 for the idealized model, at the end of cycle 73, the fifth least-squares cycle. In view of the effort that had been expended and the imminent availability of a restrained least-squares refinement program (Konnert, 1976; Hendrickson, 1976), we terminated this part of the refinement after cycle 73.* An account of continued refinement by restrained least-squares appears in the following publication (Ramanadham, Sieker & Jensen, 1990).

Discussion

The refinement of the triclinic HEW lysozyme model proved to be a more taxing problem than was originally anticipated. The difficulty was caused by the number and kinds of errors present in the initial model, complicated by the highly ordered solvent structure.

The model for triclinic lysozyme was derived from the tetragonal one both by the rotation function and by comparing Fourier transforms, the solutions being based on 6 Å resolution data (Joynson *et al.*, 1970). Models derived in this way may not be very accurate because the relative positions of parts of the molecule, particularly the side chains, may differ substantially in the parent and in the derived structure. In addition, any errors in the parent model will be inherent in the derived one.

Since we were concerned at the outset about the validity of the initial model, we restricted the data set to reflections with $d > 2$ Å in order to maximize the convergence range and omitted those with $d > 10$ Å to minimize the effects of missing solvent. R decreased rapidly at first (Figs. 1 and 2), then only slowly, reaching a value of 0·20 (unidealized) by cycle 13, and having essentially converged by cycle 17. For reasons cited earlier, we questioned the validity of the refinement against the original 2 Å data. Accordingly, we added the data beyond 2 Å, and it was immediately evident from the high R values for these data that substantial errors remained in the model (see Fig. 3, plot for cycle 17). Indeed, in the subsequent refinement against the 1·5 Å data set, numerous errors were identified by inspecting $F_o - F_c$ and $2F_o - F_c$ maps, many of such magnitude that they could be corrected only by manually adjusting the atoms or groups of atoms involved.

The side chains of arginine residues illustrate the difficulties with the initial model. In the very first $F_o - F_c$ map it was clear that some arginine side chains were seriously in error, but the extent of these was not evident until other model errors had been corrected. In fact, it was not until we were refining against the 1·5 Å data set that a check of all eleven arginine side chains revealed that seven of them still suffered sufficiently large errors that manual repositioning or rebuilding was required.

The most troubling region in the protein was in the vicinity of Pro70. As noted earlier, the electron density in the loop from Asn65 through Arg 73 was relatively low, and no reasonable density was observed for Pro70. The position of Pro70 was

* Coordinates for the idealized model at the end of cycle 73 and structure factors have been deposited with the Protein Data Bank, Brookhaven National Laboratory (Reference: 1LZT, R1LZTSF) and are available in machine-readable form from the Protein Data Bank at Brookhaven or one of the affiliated centres at Melbourne or Osaka. The data have also been deposited with the British Library Document Supply Centre as Supplementary Publication No. SUP 37029 (as microfiche). Free copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England. At the request of the authors, the list of structure factors will remain privileged until 1 December 1990.

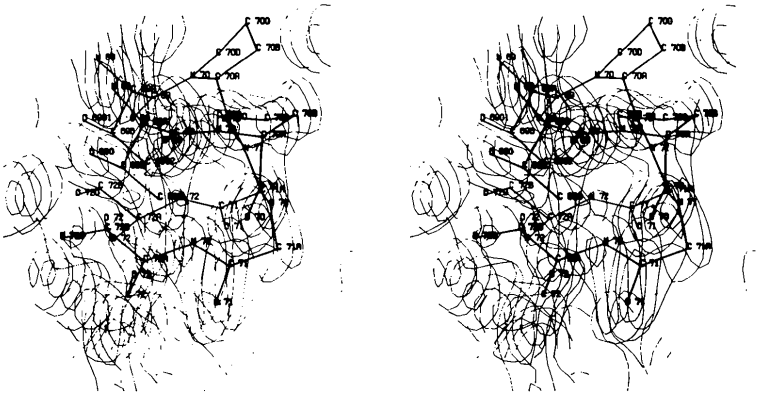

Fig. 9. Stereoview of skeletal models corresponding to initial coordinates and to the coordinates after cycle 71 with superposed electron density from a $2F_o - F_c$ synthesis based on phases from cycle 71.

evident neither in $F_o - F_c$ maps nor in $2F_o - F_c$ maps based on calculated phases omitting the residue; it finally became clear after many additional cycles, including manually adjusting nearby residues, particularly Gly71 and Ser72. Fig. 9 shows skeleton models of residues 69 through 72 corresponding to the initial coordinates and to those after cycle 71 superimposed on the electron density of a $2F_o - F_c$ map of the region. The electron density is well shaped and corresponds closely to the model after cycle 71.

Diffuse or indefinite regions of electron density in macromolecular maps are often ascribed to structual disorder. We considered this as a possible explanation of the difficulty in the region of Pro70 as long as no reasonable density appeared, but further refinement as noted above demonstrated that disorder was not the explanation; instead, the phase errors had been sufficient to obscure the residue.

The solvent structure in triclinic lysozyme is mostly ordered making it difficult to detect and correct improperly placed protein atoms, particularly those in the side chains. Although the solvent content of the triclinic crystals is relatively low, it is still 26% of structure, and since it is mostly ordered, it contributes proportionately more to the intensities than does the solvent in typical protein crystals. Thus, the solvent structure must be determined more accurately in order to insure convergence of the refinement.

The side chain of Glu7 illustrates the difficulty caused by the highly ordered solvent. We observed that the geometry of the side-chain carboxyl group became seriously distorted during cycles of free refinement between idealizations. It soon became apparent that what had been thought to be nearby solvent molecules displayed the expected shape of a glutamic acid side chain beyond C^{β} and that this density could be neatly fitted by an approximate 180° rotation of the Glu7 side chain about the C^{α}—C^{β} bond. When corrected in this way, the group behaved well in subsequent refinement cycles. The distortions we had observed in the free refinement cycles resulted from shifts of the erroneous Glu7 side-chain atoms to fit the water structure in the region.

Beginning in cycle 25, the model for both the protein and solvent was extensively edited by adding or deleting atoms or groups of atoms and by adjusting them on the basis of $F_o - F_c$ or $2F_o - F_c$ maps. It was during these cycles that some of the most troublesome errors in the model noted above were corrected, yet we never observed by dramatic decrease in R (see Fig. 2).

Beyond cycle 17 in Fig. 2, we note that when the model was idealized the increase in R was much larger for some cycles than for others. The relatively large increases occurred when large numbers of atoms were removed as in cycles 25 and 34 or when other extensive changes had been made with only a small change in the number of atoms as in cycles, 41, 68, 71 and 72. In the latter three cycles, that part of the increase in R that is solely due to the idealization is shown by the intermediate points.

When we first observed the trend in the group scale factors k' after cycle 17 (Fig. 4), we thought that by allowing the overall scale k to float downward and the mean B to adjust upward we could bring k' values for reflections with $\sin\theta/\lambda > 0{\cdot}1$ Å^{-1} satisfactorily close to unity. That this could be done was clearly established by cycle 21. As noted above, however, k' for the group of reflections with $\sin\theta/\lambda < 0{\cdot}1$ Å^{-1} was then much greater than unity. The discrepancy for the low-angle reflections suggests secondary extinction as a possible explanation, since $\sum F_o < \sum F_c$ for these reflections. The possibility was checked (Stout & Jensen, 1968), and although a few intense, low-angle reflections appeared to suffer extinction, the correction did not substantially change the distribution of k' shown in Fig. 4. Thus, no combination of adjusting the B parameters and overall scale, k, along with possible secondary extinction was found that could account for the observed distribution of k'.

The systematic deviations from unity for the ratio $I_{\overline{hkl}}/I_{hkl}$ shown in Fig. 6 are unlikely to be a significant factor in the observed distribution of k'. Reflections with $\sin\theta/\lambda < 0{\cdot}2$ Å^{-1}, corresponding to the first three points on each plot in Figs. 4 and 5, are all from crystal 1 which was free of the effect; and for crystal 2, corresponding to the fourth point in the plots, the effect was small. Only for crystals 3 and 4, corresponding to the last two points in each plot, did the ratio $I_{\overline{hkl}}/I_{hkl}$ deviate seriously from unity. If data from these crystals are properly scaled to the rest of the data, however, any effect on the value of k' for the two groups of reflections would be minimal.

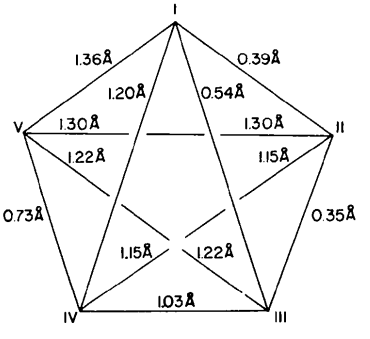

Fig. 10. Root-mean-square differences in coordinates at five stages of the refinement: (I) initial; (II) after two ΔF cycles; (III) after 15 additional differential difference cycles, 2 Å data; (IV) after 51 additional difference cycles, 1·5 Å data; (V) after five additional least-squares cycles, corrected 2 Å data.

A significant factor which will affect the distribution of k is the lack of a complete solvent model. Crystals of triclinic lysozyme have the equivalent of ~300 water molecules for each protein molecule. Although from cycles 9–25 there were 332 water sites, most of these were partially occupied and many were later deleted, leaving 220 sites by the last least-squares cycle, accounting for approximately half of the solvent molecules in the unit cell.

Fig. 10 summarizes the average r.m.s. shifts in protein coordinates for each of the five stages in the refinement. The total r.m.s. shift after 73 cycles was 1·36 Å.

Concluding remarks

One is well advised to ask why the refinement was so slow and what can be learned from it. We note the following:

(1) Models derived by the rotation function from related structures will inherit not only any errors in the related model but any differences in the molecule in the two structures. This could be as major as substantial changes in the main chain or as pervasive as differences in external side chains stemming from differences in molecular packing.

(2) Models derived by the rotation function by use of low-resolution data may be substantially improved by the use of higher-resolution data. Thus, Moult, Yonath, Traub, Smilansky, Podjarny & Saya (1976) found that in a least-squares adjustment of their triclinic lysozyme model against 2·5 Å X-ray data, one of their orientation angles changed by 4·2° from the initial value, reducing R from 0·52 to 0·418.

(3) In the early stages of our refinement against the 2 Å data, R decreased to a value that was considered acceptable at the time, yet subsequent refinement showed that numerous errors remained in the model. This emphasizes the hazard of accepting a model solely because R has been reduced to a value that appears to be in an acceptable range.

(4) The first $F_o - F_c$ map in the present refinement showed most of the solvent to be ordered. Accordingly, we added solvent to the model at an early stage (see Fig. 2). This expedited refinement in the sense of reducing R, but at the cost of unacceptable errors in the solvent.

(5) The solvent was edited repeatedly during the refinement against the 1·5 Å data, deleting and adding atoms on the basis of $F_o - F_c$ maps, and again after cycles 71 and 72 in refining against the corrected 2 Å data. In retrospect, we believe it would have been more effective to have removed the solvent from the model near the end of the refinement and then to have redetermined it, see Ramanadham *et al.* (1990).

(6) The systematic errors discovered in the data from crystals 3 and 4 undoubtedly slowed the refinement against the 1·5 Å data set, but we do not know to what extent. Despite the problem with crystals 3 and 4, some of the most serious errors in the model were corrected during the refinement against the higher-resolution data set.

We express thanks to Professor A. C. T. North for the lysozyme coordinates, to Dr K. D. Watenpaugh for the use of his blocked-least-squares program, to Professor J. Hermans for the idealization program, and to Dr R. E. Stenkamp for technical assistance. This work was supported by Grants AM-3288 and GM-10828 from the National Institutes of Health.

References

ABRAHAMS, S. C. & KEVE, E. T. (1971). *Acta Cryst.* A**27**, 157–165.
BLAKE, C. C. F., KOENIG, D. F., MAIR, G. A., NORTH, A. C. T., PHILLIPS, D. C. & SARMA, V. R. (1965). *Nature (London)*, **206**, 757–761.
BOOTH, A. D. (1946). *Trans. Faraday Soc.* **42**, 444–448; 617–619.
BUSING, W. R. & LEVY, H. (1957). *J. Chem. Phys.* **26**, 563–568.
COCHRAN, W. (1951). *Acta Cryst.* **4**, 408–411.
DIAMOND, R. (1974). *J. Mol. Biol.* **82**, 371–391.
FREER, S. T., ALDEN, R. A., CARTER, C. W. & KRAUT, J. (1975). *J. Biol. Chem.* **250**, 46–54.
HENDRICKSON, W. A. (1976). Verbal presentation at the International School of Crystallography, 'Etore Majorana' Centre for Scientifice Culture, Erice, Trapani, Italy, April 1976.
HERMANS, J. JR & MCQUEEN, J.E. JR (1974). *Acta Cryst.* A**30**, 730 739.
HODSDON, J. M., SIEKER, L. C. & JENSEN, L. H. (1975). *Am. Crystallogr. Assoc. Abstr.* **3**, 16.
JOYNSON, M. A., NORTH, A. C. T., SARMA, V. R., DICKERSON, R. E. & STEINRAUFF, L. K. (1970). *J. Mol. Biol.* **50**, 137–142.
KONNERT, J. H. (1976). *Acta Cryst.* A**32**, 614–617.
KURACHI, K., SIEKER, L. C. & JENSEN, L. H. (1976). *J. Mol. Biol.* **101**, 11–24.
LUZZATI, V. (1952). *Acta Cryst.* **5**, 802–810.
MOULT, J., YONATH, A., TRAUB, W., SMILANSKY, A., PODJARNY, A., RABINOVICH, D. & SAYA, A. (1976). *J. Mol. Biol.* **100**, 179–195.
NORTH, A. C. T., PHILLIPS, D. C. & MATHEWS, F. S. (1968). *Acta Cryst.* A**24**, 351–359.
RAMANADHAM, M., SIEKER, L. C. & JENSEN, L. H. (1990). *Acta Cryst.* B**46**, 63–69.
STEINRAUF, L. K. (1959). *Acta Cryst.* **12**, 77–79.
STOUT, G. H. & JENSEN, L. H. (1968). *X-ray Structure Determination*, 1st ed., pp. 411–412. New York: Macmillan.
WATENPAUGH, K. D., SIEKER, L. C., HERRIOTT, J. R. & JENSEN, L. H. (1972). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 359–367.
WILSON, A. J. C. (1942). *Nature (London)*, **150**, 152.