J. Mol. Biol. (1976) **100**, 179–195

The Structure of Triclinic Lysozyme at 2·5 Å Resolution

J. Moult†, A. Yonath, W. Traub, A. Smilansky,
A. Podjarny, D. Rabinovich and A. Saya‡

Departments of Structural Chemistry and Chemical Physics‡
Weizmann Institute of Science, Rehovot, Israel

(Received 2 May 1974, and in revised form 7 July 1975)

The gross similarity of the conformation of the hen egg-white lysozyme molecule in the triclinic and tetragonal crystal forms is known from an earlier study. In this work we have established the detailed conformation of the molecule in the triclinic form and compared the two structures using appropriately weighted difference maps. An independent model of the triclinic structure has been obtained by use of the real-space refinement technique. There are appreciable conformational differences of main chain, as well as side-chains, but these all occur in surface regions involved in intermolecular contacts. The locations of ordered solvent molecules have been determined.

1. Introduction

The tetragonal crystal form of hen egg-white lysozyme was the first enzyme structure to be determined at high resolution (Blake *et al.*, 1965,1967*a*; Phillips, 1967). This protein is also known to exist in several other crystal forms (Steinrauf, 1959; Imoto *et al.*, 1972). One of these is triclinic, and the molecular orientation in such crystals was determined at 6 Å resolution by the use of a rotation function and a molecular transform search (Joynson *et al.*, 1970). This work established the gross similarity of the molecular conformation in the tetragonal and triclinic crystals.

We describe here a determination of the triclinic structure at relatively high resolution (2·5 Å), and a comparison of it with the tetragonal structure. Extensive use has been made of difference maps and a refinement procedure to establish fairly precisely the differences between the triclinic crystal structure and the molecular model previously deduced from the tetragonal form. In conjunction with these, we also examined difference maps relating this model and the observed data for tetragonal lysozyme.

We undertook this investigation with several objectives in mind. First, we were interested in establishing the triclinic structure in order to study chemical complexes of lysozyme for which the tetragonal form is not suited. In particular, we have been able to prepare glutaraldehyde crosslinked triclinic crystals, complexed with denaturing agents, which are isomorphous with the native form (Yonath *et al.*, 1973). Crosslinking of tetragonal crystals, on the other hand, causes disorder which destroys their symmetry (Haas, personal communication). It may also be possible to extend studies of lysozyme–inhibitor complexes, which have provided much insight into the nature of enzyme catalysis (Blake *et al.*, 1967*b*; Beddell *et al.*, 1970; Imoto *et al.*,

† Present address: Physics Department, University of Edinburgh, Scotland.

1972), to include substrate binding at subsite E. Because of intermolecular contacts this site is not accessible to inhibitor in the tetragonal form, but it appears to be so in triclinic crystals. We were also interested in a detailed comparison of the conformations of the same protein in different environments in order to throw light on the stability of the molecular structure and the nature of the packing forces in the crystals. Finally, we sought to determine the positions of water molecules, which are often indicated in electron density maps of proteins by peaks appearing just outside the protein region. If such a stable water-coating exists around a globular protein in solution, it may well be important to the enzyme mechanism and to the stability of the molecular conformation. We were therefore interested in comparing this "water structure" in different environments to see how characteristic of the conformation it is.

A preliminary report of this work has already been presented (Moult *et al.*, 1973).

2. Data Collection

Triclinic crystals of hen egg-white lysozyme were grown according to the procedure of Steinrauf (1959). Filtered solutions of 10 mg hen egg-white lysozyme/ml (salt-free, Worthington) in 2% sodium nitrate and 0·025 M-acetate buffer, pH 4·5 were prepared. Large needles (~2 mm) of monoclinic lysozyme were usually formed, but on one occasion small (~0·1 mm) crystals of triclinic lysozyme also appeared. Use of these crystals for seeding similar solutions resulted in the growth of large triclinic crystals.

Precession photographs and diffractometry established the unit-cell dimensions as $a = 27{\cdot}4$ (5) Å, $b = 31{\cdot}9$ (9) Å, $c = 34{\cdot}4$ (9) Å, $\alpha = 88$ (·7)°, $\beta = 108$ (·6)°, $\gamma = 112$ (·0)°, which are close to the values reported by Steinrauf (1959). The data were observed to extend to a resolution beyond 1·2 Å.

Three-dimensional diffraction data to a minimum spacing of 2·5 Å were measured on a precession diffractometer (Rabinovich & Schmidt, 1973). This is an automatic two-circle diffractometer, based on cone-axis geometry, and a detailed description of this instrument will be given elsewhere (Rabinovich, manuscript in preparation). Briefly, however, the principles can be conveniently explained in terms of the precession method (Buerger, 1964). Intensities on any one plane of the reciprocal lattice are collected for a fixed setting of the instrument by precessing the normal to this plane about the incident beam in a fashion similar to that employed in the precession camera. The reflected beams of the reciprocal lattice points belonging to the selected plane emerge from the crystal in directions along the generator of a cone whose half angle $\bar{\nu}_n$ is determined by n, the order of the plane, and $\bar{\mu}$, the tilt angle between the normal to the chosen plane and the incident beam. A scintillation counter aligned along the cone generators intercepts the reflected beams and measures their intensities. The circle settings for each reflection in the reciprocal lattice plane are determined by a mini-computer.

Details of the data collection are given in Table 1. The large precession angles employed, and the non-alignment of the quartz capillary axis with the spindle axis of the instrument, resulted in appreciable absorption effects, similar to those often encountered in collecting protein data with four circle diffractometers. The usual type of semi-empirical absorption correction based on measuring the relative absorption about an axis of rotation (North *et al.*, 1968) is not appropriate to the precession

TABLE 1

Data collection

Crystal no.	Spindle orientation	R_F† factor for zero levels without absolute correction	R_F† factor for zero levels with absolute correction	Absorption correction details: μ protein (cm^{-1})	Absorption correction details: μ capillary (cm^{-1})	Levels collected	No. of reflections
1	b^*	0·06	0·05	8	160	3 levels ⊥ to 100	1374
2	c^*	0·11	0·05	12	160	6 levels ⊥ to 010	1774
		0·09	0·03	12	160	5 levels ⊥ to 110	1687
3	$-c^*$	0·05	0·03	8	160	6 levels ⊥ to 110	1410
		0·08	0·02	8	160	7 levels ⊥ to 120	1859
						Total	8104

A total of 8104 reflections were collected. After rejecting 400 intensity measurements with values less than 4 times the standard deviation and scaling by the method of Rollett & Sparks (1960), these reduced to 3423 reflections in the unique set.

Overall R factor defined as

$$R = \frac{\sum_{\text{all } hkl} w_{jl}|\bar{F} - k_j F_{jl}|}{\sum_{\text{all } hkl} w_{jl} k_j F_{jl}} = 0{\cdot}0425,$$

where w_{jl} is the weight for each F_{jl} derived from the counting statistics, and k_j is the scale factor for level j.

† R_F factor for Friedel pairs of zero levels calculated as

$$R_F = \frac{\sum(F^2 - F_i^2)}{\sum F_i^2}.$$

geometry. Therefore an integrated absorption correction with calculated absorption paths through both crystal and capillary was used (Coppens *et al.*, 1965). The presence of mother liquor around the crystal was allowed for by adjustment of the absorption parameters, μ, for the crystal and capillary tube. The marked improvement in the agreement factors between Friedel pairs of zero level reflections resulting from this correction is shown in Table 1. After correction for absorption the data were also corrected for Lorentz and polarization factors and all the levels scaled together using the method of Rollett & Sparks (1960), with weights derived from counting statistics in the usual way (Arndt & Willis, 1966).

3. Refinement of the Molecular Orientation

A set of structure factors out to 2·5 Å resolution was calculated for the molecule positioned at the orientation given by Joynson *et al.* (1970). These structure factors correspond to atomic co-ordinates for the tetragonal structure (Imoto *et al.*, 1972) which had been made stereochemically reasonable by Diamond's model-building procedure (Diamond, 1966) and are referred to by him as set M2 (Diamond, 1974). The observed structure factors were scaled to the calculated values, using a Wilson

plot, and a temperature factor (B) of 8, chosen as giving the best agreement for all shells of $\sin\theta/\lambda$.

A conventional unweighted R factor

$$\left(R = \frac{\sum||k|F_o|-|F_c||}{\sum k|F_o|}\right)$$

between the observed and calculated structure factors was found to be 52%, for 2·5 Å data (44·2% for 6 Å), with the scale factor k equal to 1·12 to allow for an estimated 100 to 200 ordered water molecules. Structure factors calculated for randomly chosen molecular orientations give R factors of approximately 60%. B and k were slightly readjusted to minimize R values at later stages of structure refinement.

This initial molecular orientation was improved by a systematic local angular search for the orientation having the lowest R value and also by a least-squares refinement based on minimization of the residual $\sum(|F_o|-|F_c|)^2$ as a function of the three rotation angles. A modification of the Eulerian angles described by Rossmann & Blow (1962) was used, with rotations first about the x axis, then about the new position of the z axis (i.e. after rotation about x) and finally about the new y axis. The choice of y for the third rotation axis, instead of the more usual x, removed the degeneracy for $\theta_2 = 0$ which can be an inconvenience in angular refinement procedures. The x and z axes of the right-handed Cartesian co-ordinates correspond, respectively, to the a and c axes of the tetragonal cell.

The results of the search and the refinement are given in Table 2. Both methods of angular refinement led to essentially the same minimum at ($\theta_1 = -47{\cdot}5°$; $\theta_2 = 4{\cdot}9°$; $\theta_3 = -6{\cdot}2°$), that is ($-0{\cdot}3°$; $1{\cdot}6°$; $-4{\cdot}2°$) away from the initial orientation. This gave an R factor of 41·8%.

Subsequently a new set of atomic co-ordinates for the tetragonal structure (RS5D), resulting from a real-space refinement (Diamond, 1974) plus some energy refinement (Levitt, 1974), became available. This gave an R factor of 40·8% at the above refined orientation and this was slightly improved to 40·3% by a further five rounds of orientation refinement to ($\theta_1 = -47{\cdot}5°$; $\theta_2 = 4{\cdot}8°$; $\theta_3 = -6{\cdot}4°$) after shifts of ($0°$; $0{\cdot}1°$; $-0{\cdot}2°$) in the three angles from the previous orientation.

4. Structure Analysis and Refinement

Two pairs of maps were employed. For the triclinic form we produced a weighted ($|F_o|-|F_c|$) difference map and an F_c map, using calculated phases for both. For the tetragonal form we used data kindly supplied by Professor D. C. Phillips to prepare a replica of the original F_o map, with phases derived from three isomorphous derivatives and weights based on the figures of merit (Blake *et al.*, 1967*a*), as well as an ($|F_o|-|F_c|$) difference map with calculated phases and the same weighting scheme as for the triclinic difference map. Both sets of maps were sectioned in the same direction relative to the molecular orientation to facilitate comparison of the two structures. Initially the difference maps were based on structure amplitudes and phases calculated from M2 co-ordinates. Subsequently, the tetragonal map was replaced by one, based upon RS5D co-ordinates, and the triclinic map by one based on the TRIC3 set of refined atomic co-ordinates (see below). Co-ordinates were included for all protein atoms, but not solvent, in both crystal forms. In both cases the use

TABLE 2

Molecular orientation refinement

(a) By systematic search (model derived from M2 co-ordinates)

Resolution (Å)	No. of reflections	Angular increment	Starting orientation $\theta_1°$	$\theta_2°$	$\theta_3°$	R† at starting orientation (%)	Final orientation $\theta_1°$	$\theta_2°$	$\theta_3°$	R† at final orientation (%)
6·0	192	5	−47·8	6·5	−2·0	44·2	−45·3	4·0	−9·5	40·2
4·0	793	5	−45·3	4·0	−9·5	41·7	−45·3	4·0	−9·5	41·7
2·5	232 strongest reflections	2·5	−45·3	4·0	−9·5		−47·8	5·0	−7·0	31·9

(b) By least-squares minimization

Resolution (Å)	Co-ordinate set	No. of reflections	Starting orientation $\theta_1°$	$\theta_2°$	$\theta_3°$	R† at starting orientation (%)	No. of cycles	Final orientation $\theta_1°$	$\theta_2°$	$\theta_3°$	R† at final orientation (%)
4·0	M2	793	−47·80	6·50	−2·00	46·5	3	−47·80	4·00	−4·01	42·5
6·0	M2	192	−47·80	4·00	−4·01	41·9	6	−47·30	4·74	−5·58	38·4
2·5	M2	3423	−47·30	4·74	−5·58	43·3	5	−47·55	4·90	−6·18	41·8
2·5	RS5D	3423	−47·55	4·90	−6·18	40·8	5	−47·54	4·79	−6·45	40·3

† $R = \dfrac{\Sigma(||F_o| - |F_c||)}{\Sigma|F_o|}$.

of the refined co-ordinates produced a marked improvement in the quality of the map.

Three different weighting schemes for the difference maps were tried. The first maps were not weighted at all. Then weighting factors of $F_c^4/(F_c^4 + F_o^4)$, which reduce the contributions from poorly determined difference vectors (Stout & Jensen, 1968), were applied. Finally, weights of

$$\operatorname{cotanh}\left(\frac{\sum|F_o||F_c|}{\sigma_Q} - \frac{\sigma_Q}{\sum|F_o||F_c|}\right)$$

(where σ_Q is the mean square length of the difference vector at that value of $\sin\theta$) were used to produce maps with the minimum root-mean-square error contribution from each reflection (Moult, Saya & Lifson, unpublished results). Use of these weighting schemes produced substantial improvements in the interpretability of the difference maps.

The triclinic maps, drawn in different colours on the same sheets, were suspended in a Richards optical comparator (Richards, 1968) for comparison with a skeletal model of the protein. The two tetragonal maps were drawn in different colours and superimposed, with appropriate separation between sections, in an illuminated viewing box.

Both difference maps show some side-chains and, particularly in the triclinic map, short segments of main chain in negative regions. In most cases these can be satisfactorily repositioned in or near neighboring positive regions (see Figs 1 and 2). Several such side-chains, in the tetragonal structure, occur in low density regions of the F_{obs} map, and had been included in the difference maps with arbitrarily assigned conformations (Diamond, 1974). There is also a large number of features indicating small errors in positions.

In order to produce an independent set of triclinic co-ordinates, some refinement away from the tetragonal model structure was undertaken. A difference map refinement procedure was developed utilising the Diamond real-space refinement program (Diamond, 1974). For this purpose we used an "observed" triclinic map, formed by the addition of the calculated map and a scaled difference map. It was found advantageous to use different scale factors for positive and negative regions of the difference map, and the best results were obtained using 2·0 for positive regions and 1·8 for negative regions of a map prepared with the "minimum error" weighting scheme described above. On theoretical grounds we might expect a scale factor of about 2·0 to be appropriate (Luzzati, 1953; Henderson & Moffat, 1971). A smaller scale factor was used for negative regions because of the relative enhancement of these by the weighting scheme. This observed map was put into the real-space refinement programme.

R factors were calculated for the co-ordinates produced by the refinement procedure using various program parameters. Attempts to refine occupancy were unsuccessful and this was fixed at 1·0 for all protein atoms, but positional refinement of three residues at a time with 10^{-3} filtering reduced the *R* factor to 36·3%. Further real-space refinements, using a map based on structure factors calculated from these new co-ordinates and the above parameters, reduced the *R* factor to 35·7% with a root-mean-square shift of atomic positions of 0·5 Å from the RS5D tetragonal co-ordinates.

Inspection of the difference map produced from these co-ordinates showed that

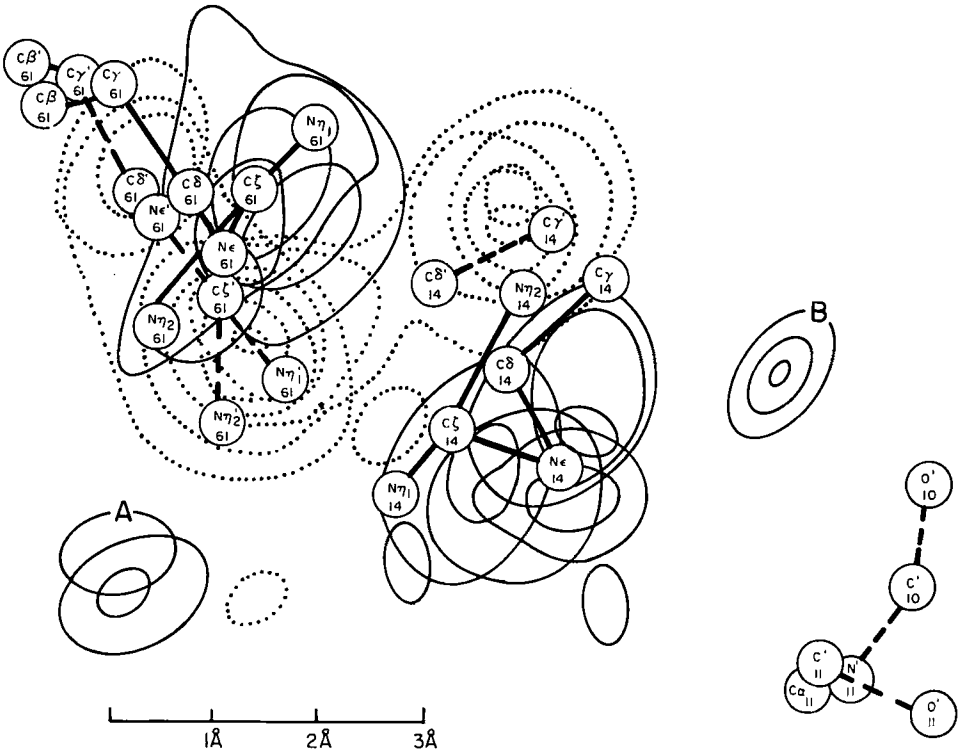

FIG. 1. Superposition of 8 sections of the triclinic difference map indicating shifts from the tetragonal co-ordinates in the positions of residues Arg61 and Arg14 in neighbouring molecules, which relieve short contacts between them. Positive and negative contours are shown as solid and dotted lines, respectively, and correspond to intervals of 0·33 e/Å^3 beginning at 0·45 e/Å^3. The M2 co-ordinates, from which calculated structure factors were obtained, are indicated as primed positions (e.g. Cγ'14) and joined by dashed lines, whereas positions refined with the aid of the difference map are shown without primes (e.g. Cγ14) and joined by full lines. Part of the main chain at residues 10 and 11 is also shown as an example of features which fall in clear regions of the difference map and whose positions were unchanged. The positions of Cβ and Cγ of both Arg61 and Arg14 were improved by the real-space refinement program, whereas atoms from Cδ onwards were first adjusted manually. Peaks A and B indicate ions or water molecules. They also appear in a difference map calculated from TRIC3 co-ordinates in which the region around Arg14 and Arg61 is clean.

shifts up to about 0·8 Å were produced successfully, but larger shifts were not, and manual intervention was therefore necessary for the inclusion of these. Accordingly, new positions for the side-chains Arg14, Arg61, Arg73, Arg125 and Arg128 were estimated from the difference map and put into a further round of real-space refinement. This gave a set of co-ordinates, TRIC3, corresponding to an R factor of 35·0%. Additional iterations of the entire refinement procedure, using various scale factors, filtering and lengths of chain segments refined at one time, failed to reduce the R factor further. The final list of TRIC3 co-ordinates is to be deposited with the Protein Data Bank.

In order to test the superiority of the combination of maps used over the more obvious choice, an attempt was also made to refine the map obtained from the observed triclinic structure amplitudes and phases derived from the rotated RS5D

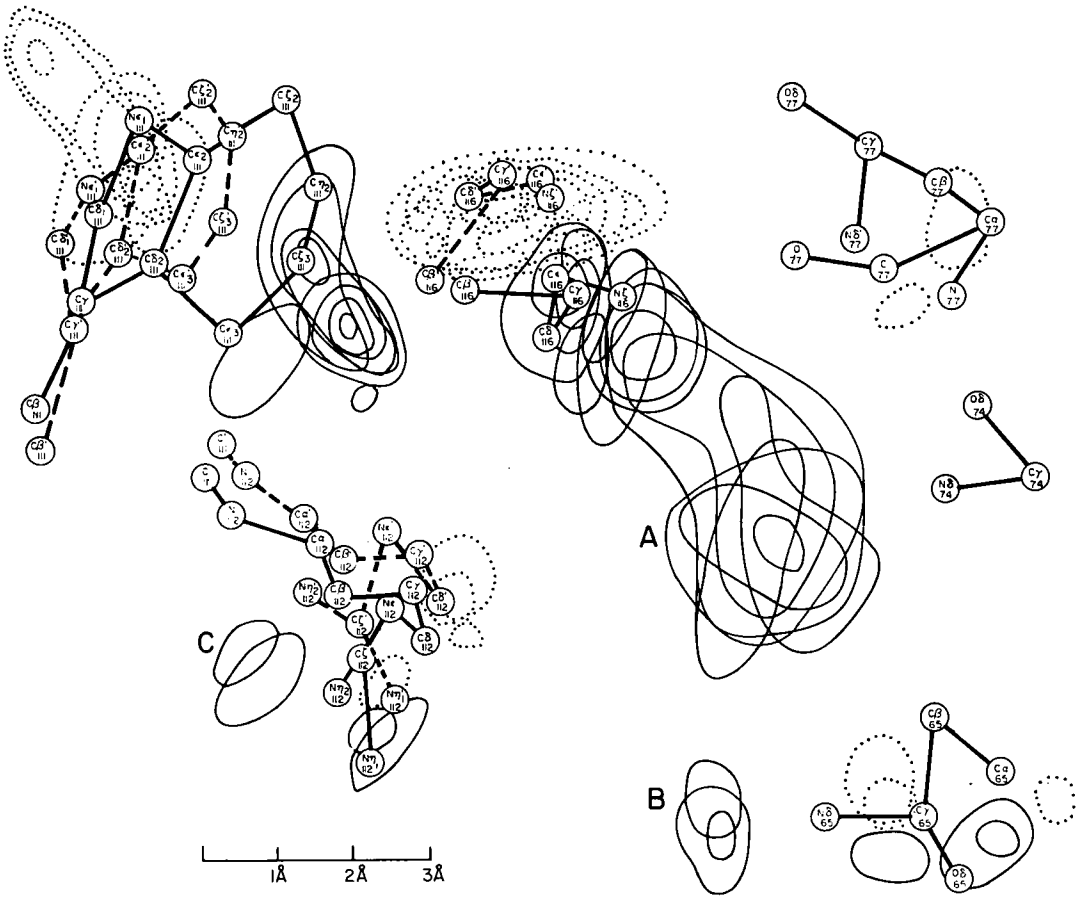

FIG. 2. Superposition of 10 sections of the triclinic difference map indicating shifts from the tetragonal co-ordinates in the positions of Lys116 and Trp111. Lys116 is involved in several intermolecular contacts (cf. Table 4) and a salt bridge with the NO_3^- counter ion (peak A in Fig.). Lys116 shields Trp111 from the solution in both structures and their positions therefore apparently change in conjunction. Positive and negative contours are shown as solid and dotted lines, respectively, and correspond to intervals of 0·33 e/Å^3, beginning at 0·45 e/Å^3. The M2 co-ordinates, from which structure factors were obtained, are indicated as primed positions (e.g. Cδ′116) and joined by dashed lines, whereas positions refined with the aid of the difference map by the real-space refinement programme are shown without primes (e.g. Cδ116) and joined by full lines. Parts of residues, Asn74 and Asn77, which are involved in the intermolecular contacts, directly or through the salt bridge, are also shown. They fall in a clean region of the difference map and their positions were almost unchanged after refinement. Asn 65 is also involved in the same kind of contacts, but falls in a fairly noisy part of the map and was still not very well-defined after the real-space refinement. Arg112 in the triclinic structure differs from the assigned tetragonal conformation only slightly at the guanidinium end, and the 2 positions are shown in the Figure. Peak A represents a NO_3^- ion and peaks B and C probably indicate water positions. These 3 peaks all appear in the difference map calculated with the TRIC3 co-ordinates.

Note that the positive contours near the tetragonal positions of Lys116 and Trp111 only indicate the directions in which these residues were moved by the refinement and not their final positions. The TRIC3 difference map is completely clean in the neighbourhood of these residues.

co-ordinates. The same real-space refinement program, with 10^{-3} filtering, was used in successive cycles. After the first one the R factor dropped to 39·4%, and after three more it fell to 38·8%. It would therefore appear that the use of the sum of a calculated map and a scaled difference map increases the power of the real-space refinement procedure. We have since learned that a similar difference map real-space refinement procedure has been used very successfully for the refinement of pancreatic trypsin inhibitor (Deisenhofer & Steigemann, 1975) and the power of difference maps for protein structure refinement, though used in a different way, has also been demonstrated in the case of rubredoxin (Watenpaugh *et al.*, 1973).

After refinement of the triclinic structure, and investigation of some uncertain features in the tetragonal structure, a detailed comparison of the two was made on a residue-by-residue basis. A list of the principal differences between the conformations is given in Table 3.

Both difference maps were searched systematically, using a computer program to produce a list of co-ordinates of all peaks with heights greater than 2·2 standard deviations. A standard deviation was calculated for the difference maps as

$$\left(\sum_{i=0}^{N} \rho_i^2/N\right)^{\frac{1}{2}},$$

where ρ_i is the density at a grid point, and N is the number of grid points in the map. This method tends to overestimate the standard deviation if there is a large number of meaningful peaks, so instead of using the common 3σ cut-off we made an empirical choice of $2{\cdot}2\sigma$, which led to the lowest R value.

Those peaks, not accounted for by shifts of the protein, were examined in terms of their environments to see if they might be water molecules. All but three were found to be reasonably located, and inclusion of these solvent co-ordinates, with occupancies estimated from peak heights, in a new calculation of structure factors for the triclinic form reduced the R value from 35·0% to 33·8%.

The tetragonal F_{obs} map shows a number of peaks in the solvent region which do not correspond to peaks in the difference map. To test their significance, the co-ordinates of 30 such peaks chosen at random were incorporated in the list of RS5D atomic co-ordinates as water molecules and a new difference map produced from these. This map had holes at all positions of the added peaks, indicating that these features in the F_{obs} map were in fact spurious. This conclusion was also supported by the fact that inclusion of the additional water positions increased the R factor from 34·9% to 36·3%.

5. Results and Discussion

The basis of this work has been the comparison of similar structures using a difference map calculated from an imperfect model of one of them. This technique has been very successful in establishing the detailed conformational differences between the two structures, although there would clearly be a level of dissimilarity between the structures and the model beyond which the map would be uninterpretable. Even the quite modest improvements in the model resulting from the refinement of the two structures causes a very marked improvement in the quality of the maps. At this level of disagreement between structure and model a weighting scheme can also do much to improve the quality of the difference maps.

The difference map for tetragonal lysozyme has served to clarify some aspects of

TABLE 3

Conformational differences between triclinic and tetragonal forms of lysozyme

Amino acid	Description	Intermolecular interactions Triclinic	Tetragonal
Arg5	Conformation of guanidinium group somewhat different in the two crystal forms, though intramolecularly hydrogen-bonded to Trp123 CO and Arg125 CO in both cases. In triclinic structure, guanidinium group also connected by several solvent bridges to 101–103 region.		
Lys13	Makes intramolecular salt bridge to terminal carboxyl in both crystal forms, but more extended side-chain in tetragonal form also allows intermolecular interaction with terminal carboxyl of dyad-related molecule. In triclinic form, side-chain curled around making good intramolecular hydrophobic contacts with Leu25 and Leu129.		Leu129 (60)
Arg14	Triclinic side-chain close to Ala10 and Ala11 with hydrogen-bond from N_ε to Ala10 CO. Differs markedly from tetragonal conformation which would imply short intermolecular contacts with Arg61 (see Fig. 1).	Asp48 (10) Arg61 (11) Asn37 (13)	Ala10 (56) Arg128 (57)
Lys33	End of side-chain in triclinic form closer to Trp123. Though this residue makes no direct intermolecular contacts in either crystal form, the terminal $N\zeta$ participates in intermolecular solvent bridges in both cases.		
Phe38	Side-chain in triclinic structure twisted by some 30° about C_α—C_β and C_β—C_γ bonds relative to tetragonal conformation, thus avoiding short intermolecular contacts with Arg73.	Arg73 (30)	
Gln41	In tetragonal form this residue lies in the dimer contact region and participates in many intermolecular contacts and solvent bridges. In triclinic form the terminal amide is oriented differently with an intermolecular hydrogen-bond from N_ε to Gln121.	Gln121 (19)	Asn65 (47) Asp66 (49) Cys80 (53) Ser81 (54)
Asn44 to Ser50	This region, which is part of a β loop, has a different main-chain conformation in the two crystal forms. In both structures there are several, but different, intermolecular contacts, and they also differ in some of their intramolecular hydrogen-bonding; whereas the tetragonal form has an Asn46 NH to Ser50 CO connection, the triclinic has hydrogen-bonds from Thr47 CO to Gly49 NH and from Gly49 CO to Thr69 O_γ.	Glu7 (8) Glu7 (9) Arg14 (10) Asn77 (17) Asn77 (18)	Gly126 (43) Cys127 (44) Arg128 (45) Asn44 to Arg45 (46) Arg68 (52)
Arg61	Whereas in the tetragonal form this side-chain is fairly extended and makes no intermolecular contacts, in the triclinic structure it is folded back with its guanidinium group interacting intermolecularly with Arg14 (see Fig. 1) and intramolecularly with O_δ of Asp48 and N_δ of Asn59.	Arg14 (11)	

TABLE 3—*continued*

Amino acid	Description	Intermolecular interactions Triclinic	Tetragonal
Arg73	Different side-chain conformations in the two crystal structures. In the triclinic form the side-chain makes many intermolecular contacts and differs from a standard extended conformation by a rotation of some 90° about the C_α—C_β bond.	Val2 (27) Phe3 (28) Asn37 (29) Phe38 (30)	Asn37 (61)
Trp111	Somewhat different conformations in tetragonal and triclinic forms. Side-chain has intramolecular contacts with Lys116, which shields it from solvent, in both structures (see Fig. 2).		
Asn113	Different intermolecular contacts and different side-chain conformations in the two structures. Closer to Arg114 side-chain in triclinic form.	Pro79 (5) Ser81 (6)	Gly102 (37) Asn103 (38) Gly104 (39) Asn106 (40) Lys116 (41)
Arg114	Different intermolecular contacts and different side-chain conformations in the two structures. Closer to Phe34 and Cys115 in triclinic form.	Gly16 (14) Asp18 (15) Tyr20 (16)	Gly22 (34) Tyr23 (35)
Lys116	Different intermolecular contacts and different side-chain conformations in the two structures. Shields Trp111 from solvent, and is hydrogen-bonded from N to Asn106 N_δ in both forms, but closer to Arg112 in triclinic form (see Fig. 2).	Asn77 (1) Ilu78 (2)	Asn113 (41)
Arg128	Different intermolecular contacts and very different side-chain conformations in the two crystal forms. In triclinic structure side-chain unextended with hydrogen bonds from $N\eta$ to Gly126 CO and Cys6 S_γ.	Val109 (12) Asp101 (33)	Thr47 (45) Arg14 (57) His15 (58) Gly16 (59)

The intermolecular interactions are described in Tables 4 (triclinic) and 5 (tetragonal) with the serial numbers given here in parentheses.

We have used the standard IUPAC-IUB (1970) notation for amino acids and individual atoms.

We have not included in this table several conformational features which differ appreciably in the co-ordinates reported for the two crystal forms, but are not well-defined in one or other structure as judged by the electron-density difference maps. Notable among these are the regions of main chain around Pro70–Ser72 and Asp101–Asn103 and the side-chains of Arg21, Arg68 and Arg125.

As we are unable to distinguish between terminal O and N atoms in side-chains of Asn and Gln, we have designated them arbitrarily in Tables 3, 4 and 5.

Reference to stereodiagrams of the tetragonal structure (Imoto *et al.*, 1972) may be helpful in visualizing the conformational differences described in this Table.

the structure. These include the locations of solvent molecules and the conformations of several side-chains, most of which were not well defined in the F_{obs} map (Imoto *et al.*, 1972) and had consequently been refined from arbitrarily assigned standard conformations (Diamond, 1974). Some of these, including Asp18 and Arg112, lie in clean regions of the difference map and are evidently close to the assigned conformations. Some others lie partly in strongly negative regions and apparently require further refinement. This would be outside the scope of this work, but in comparing the triclinic and tetragonal conformations (see Table 3), we have noted such uncertain features and also considered alternative conformations for Arg73, Arg128 and the guanidinium group of Arg61 which fit into positive regions of the tetragonal difference map.

A difference map for triclinic lysozyme was used (as described in Structure Analysis and Refinement) for the real-space refinement procedure, which has been developed for refinement against F_{obs} maps (Diamond, 1974). This appears to be a most promising way of refining protein structures without being tied to phases derived from the isomorphous derivatives.

It has been estimated that the real-space refinement of lysozyme has led to co-ordinates of well-defined features having an accuracy of better than 0·2 Å (Diamond, 1974). This should be true of much of the molecule, but there are several regions, corresponding to unclear portions of the difference maps, where the errors may be appreciably larger, and for a few poorly defined side-chains, errors in some co-ordinates probably exceed 1 Å. We have compared the accuracy of the two structure determinations by calculating the normalized R factor:

$$R_n = \frac{\sum||F_o|/|\bar{F}_o| - |F_c|/|\bar{F}_c||}{\sum|F_o|/|\bar{F}_o|},$$

where $|\bar{F}_o|$ and $|\bar{F}_c|$ are averaged values for each range of $\sin\theta$ (Srinivasan & Ramachandran, 1965). R_n is 36% for the triclinic structure and 34% for the tetragonal, both at 2·5 Å resolution, and the R_n values show similar variations with $\sin\theta$, which indicates that the two structures are similar regarding distribution of errors and omission of scattering matter (Moult, Saya & Lifson, unpublished data).

Analyses of the maps have shown that the conformations of the lysozyme molecule in the two crystal forms are very similar, and, in fact, the root-mean-square difference in co-ordinates of corresponding backbone atoms is only 0·5 Å. We have found no appreciable differences in the internal portion of the molecules, but there are considerable conformational variations, involving regions of main chain as well as side-chains, on the surface. These apparently arise from the different modes of packing in the two crystal forms causing different regions of the protein to be involved in intermolecular contacts.

Table 3 describes the differences in conformation between the two crystal forms and the location of intermolecular contacts. Intramolecular interactions are also mentioned in some cases, particularly where they differ in the two forms, but for the most part we have not repeated the extensive description of these which have already been reported (Blake *et al.*, 1967*a*; Browne *et al.*, 1969; Imoto *et al.*, 1972). More details of the intermolecular contacts are given in Tables 4 and 5.

The lysozyme molecules are more closely packed in the triclinic than in the tetragonal crystals (Steinrauf, 1959), and there are more intermolecular contacts in the former (cf. Tables 4 and 5). This probably accounts for the apparent temperature

TABLE 4

Intermolecular contacts in triclinic lysozyme

Ref. no.	Molecule A atoms		Distances (Å)
		Molecule B atoms ($x+1, y+1, z$)	
1	Asn77 O	Lys116 Cγ, Cε	3·7, 3·5
2	Ilu78 Cδ1	Lys116 O	3·6
3	Ilu78 Cδ1	Gly117 C, O, Cα	3·1–3·7
4	Pro79 Cγ	Arg112 O	3·6
5	Pro79 Cβ, Cγ	Asn113, C, O	3·3–3·8
6	Ser81 Oγ	Asn113, O	3·7
7	Ala82 Cβ	Thr118 Cγ2	3·7
		Molecule C atoms ($x+1, y, z+1$)	
8	Glu7 Cβ, Cγ, Cδ, Oε1, Oε2	Thr47 C, Cα, Cβ, Oγ1, Cγ2	2·9–4.0
9	Glu7 Oε2	Asp48 N	3·6
10	Arg14 Cζ, Nη1, Nη2	Asp48 O, Cβ, Cγ, Nδ2	2·4–3·9
11	Arg14 Cβ, Cγ, Cδ, Nε, Cζ, Nη1, Nη2	Arg61 Cδ, Nε, Cζ, Nη1, Nη2	2·7–3·8
12	Arg128 Cγ	Val109 Cγ2	3·8
		Molecule D atoms ($x+1, y, z$)	
13	Arg14 O	Asn37 Nδ2	3·5
14	Gly16 O	Arg114 Cζ, Nη1, Nη2	2·9–3·2
15	Asp18 O	Arg114 Nη2	3·1
16	Tyr20 Cα, Cβ, Cγ, Cδ1, Cε1	Arg114 Cζ, Nη1	2·8–3·8
17	Asn77 Nδ2	Arg45 O, Cγ	3·0, 2·9
18	Asn77 Nδ2	Asn46 O	3·5
		Molecule E atoms ($x\ y+1, z$)	
19	Gln41 Nε1	Gln121 Nε1	2·8
20	Asp66 C, O, Cα	Arg21 Cζ, Nη2	2·2–3·7
21	Arg68 Cζ, Nη1, Nη2	Tyr23 Cζ, Oη	2·6–3·8
22	Arg68 Nη1	Asn103 O	3·7
23	Arg68 Cζ	Asn106 Nδ2	3·6
24	Ser81 O, Cβ	Asn19 Nδ1	3·3, 3·7
25	Ser81 Cβ	Gly22 Cα	3·9
26	Leu84 Cβ	Asn19 Nδ1	3·6
		Molecule F atoms ($x, y, z+1$)	
27	Val2 O, Cγ1	Arg73 Cγ, Cδ, Nε, Cζ	2·7–3·6
28	Phe3 C, O	Arg73 Cδ, Nε, Cζ, Nη1, Nη2	2·2–3·7
29	Asn37 O	Arg73 Nη2	3·2
30	Phe38 Cβ, Cγ, Cδ2, Cε2	Arg73 Cζ, Nη1, Nη2	2·0–3·8
31	Arg125 O	Asp101 O	3·5
32	Gly126 C, Cα	Asp101 C, O	2·6–3·8
33	Arg128 Cζ, Nη1, Nη2	Asp101 Cα, Oδ	3·2–3·7

The Table lists all intermolecular carbon–carbon contacts up to 4·0 Å and contacts between other atoms up to 3·7 Å. The contacts are grouped according to the pairs of molecules which they connect. Locations of residues Arg21, Arg45, Arg68, Arg73, Asp101 and Gln121 are not well-defined.

As we are unable to distinguish between terminal O and N atoms in side-chains of Asn and Gln, we have designated them arbitrarily in Tables 3, 4 and 5.

TABLE 5

Intermolecular contacts in tetragonal lysozyme

Ref. no.	Molecule A atoms		Distances (Å)
		Molecule B atoms (related by 4_3 axis)	
34	Gly22 C, O	Arg114 Nη1	3·8, 3·0
35	Tyr23 Cα, Cγ, Cδ1, Cε1, Cζ, Oη	Arg114 Cδ, Nε, Cζ, Nη1	3·0–3·8
36	Gly102 O	Arg112 Nη2	3·7
37	Gly102 O	Asn113 Cγ, Nδ1, Nδ2	3·0–3·7
38	Asn103 C, O, Cα	Asn113 Cγ, Nδ1, Nδ2	3·3–3·8
39	Gly104 N	Asn113 Nδ2	3·3
40	Asn106 Cγ, Nδ2	Asn113 C, O, Cα, Cβ	2·7–3·6
41	Lys116 Cε, Nζ	Asn113 O	3·3, 2·9
		Molecule C atoms (related by 2-fold screw axis)	
42	Gln121 Cδ, Nη2	Trp62 Cε1	3·4, 4·0
43	Gly126 O	Thr47 Oγ	3·1
44	Cys127 O	Asp48 Cα, Cβ	3·7, 3·3
45	Arg128 Cγ, Cδ, Nε	Thr47 C, O	3·0–3·9
		Molecule D atoms (related by 2-fold rotation axis)	
46	Arg45 Nη1	Asn44 Nδ1	3·7
47	Asn65 C, O	Gln41 Cδ, Nε1, Nε2	3·0–3·7
48	Asp66 O	Asn39 Nδ2	2·9
49	Asp66 C, O, Cα	Gln41 Cβ, Cδ, Nε1	3·0–4·0
50	Gly67 C, O	Asn39 Nδ2	3·6, 3·7
51	Arg68 Cζ, Nη1, Nη2	Thr43 C, O, Oγ1	2·9–3·4
52	Arg68 Nη1	Asn44 Nδ1	3·6
53	Cys80 N, C, Cα, Cβ	Gln41 Nε2	3·3–3·7
54	Ser81 N, Cα, Cβ, Oγ	Gln41 Cδ, Nε2	2·8–4·0
55	Leu84 Cδ1	Leu84 Cδ1	3·9
		Molecule E atoms (related by 2-fold rotation axis)	
56	Arg14 Cγ, Cδ	Ala10 Cβ	3·3, 3·6
57	Arg128 Cδ	Arg14 O	3·5
58	Arg128 Cζ	His15 C	3·3
59	Arg128 Cζ, Nη1	Gly16 N, Cα, C	2·3–3·3
60	Leu129 C, O	Lys13 Nζ	3·4, 3·4
		Molecule F atoms (related by translation along c axis)	
61	Asn37 Nδ2	Arg73 Nη2	3·2

The Table lists all intermolecular carbon–carbon contacts up to 4·0 Å and contacts between other atoms up to 3·7 Å. The contacts are grouped according to the pairs of molecules which they connect. Intermolecular distances were calculated from the RS5D set of tetragonal co-ordinates (Diamond, 1974). The locations of residues Trp62, Arg73, Gln121 and Arg128 are not well-defined.

As we are unable to distinguish between terminal O and N atoms in side-chains of Asn and Gln, we have designated them arbitrarily in Tables 3, 4 and 5.

factor being lower in the triclinic ($B = 8$) than in the tetragonal form ($B = 15$). The strong tetragonal "dimer" interactions, which it was suggested might be maintained in solution (Blake *et al.*, 1967*a*), do not occur in the triclinic form. In fact, there are no intermolecular interactions common to the two structures, though mole-

cules separated by a translation along the *c* axis have nearly the same contacts between Arg73 and Asn37.

In several cases, long flexible side-chains have very different conformations in the two structures, and the maximum discrepancies in positions of corresponding atoms in Arg14, Lys33, Phe38, Arg61, Arg73, Arg114 and Arg128 range from 2 Å to 7 Å. There are also some differences in main-chain conformations, notably in the β-loop region between residues 44 and 50 where several atomic positions differ by almost 2 Å.

Such differences can often be directly related to different intermolecular contacts (see Table 3), but sometimes the reasons may be quite indirect. Thus, for example, the different intermolecular contacts made by Lys116 affect not only its own conformation, but also that of the hydrophobic Trp111 which it protects from the solvent (cf. Fig. 2). All of the residues that constitute the substrate binding site of lysozyme, with the possible exception of Asp101, appear to have the same conformation in the two crystal forms. Apart from Trp62 there are no disordered side-chains in the triclinic structure, though there are several in the tetragonal form (Blake *et al.*, 1967*a*).

In both crystal forms about a third of the total solvent molecules (as estimated from unit-cell dimensions and partial specific volume of the protein) were located in the difference maps. Of approximately 140 solvent molecules found in the tetragonal structure and some 110 in the triclinic, about 60 occur at similar locations on the protein in the two crystal forms, that is they have at least two identical protein ligands. These include the water molecules at the active site of lysozyme (Phillips, 1971), as well as three internal water molecules (Imoto *et al.*, 1972). Almost all the solvent molecules have three ligands and, almost invariably, at least two of these are from the lysozyme molecule. Thus, apart from a few small regions (e.g. near Lys13 and Lys33), there is no second shell of solvent structure apparent in our difference maps.

Practically all the accessible polar and charged groups interact with solvent. Of these about 180 are common to the two forms and most of the others are precluded from solvent interactions in one or other of the structures by intermolecular contacts.

The two crystal forms have different protein–protein contact points, at which there are almost no bound water molecules. Consequently, we were able to merge the two "water-shells" together and extrapolate from them a water structure around the protein which is not affected by crystal packing, and may represent the mode of water binding in solution.

We should point out that the difference electron density chosen to identify solvent peaks, and hence their number, is somewhat arbitrary, as they may have appreciably less than 100% occupancy. Indeed, since this work was completed, we have learned of an analysis of solvent in the tetragonal structure by somewhat different methods which indicated a similar but not identical number of solvent positions (D. C. Phillips, personal communication).

In both structures there are several particularly large solvent peaks near positively charged side-chains which are probably due to bound anions. Some of these, near Arg21, Arg112, Arg114 and Lys116, have similar locations in the two structures. The NO_3^- anions of the triclinic crystals are stronger hydrogen-bond acceptors than the Cl^- ions of the tetragonal form, and the difference maps indicate that they also bind to several hydrogen-bond donors in the protein which are not near positively-

charged side-chains. This may account for the different intermolecular interactions which determine the two crystal forms.

The results of this investigation have some general implications concerning the use of X-ray studies in elucidating the biological mechanisms of proteins. There is much evidence for the uniqueness of protein tertiary structure under functional conditions (Blow & Steitz, 1970) and, consequently, there has been a tendency to assume that crystal structures determined by X-ray diffraction correspond to biologically active conformations. However, there have been few detailed tests of the effects of crystalline environment on protein conformation. This study indicates that, though the structure of the interior of the molecule is quite stable, there is considerable variability in the surface structure. Similar results have been reported recently from comparative conformational investigations of subtilisin (Drenth *et al.*, 1971) and chymotrypsin (Tulinsky *et al.*, 1973). The effects of pH have been highlighted in one of these studies (Vandlen & Tulinsky, 1973), whereas the conformational variations we have observed can perhaps be ascribed entirely to crystal packing forces, though we cannot rule out the possibility that some changes may be caused directly by protein–ion interactions. These forces can evidently affect significantly the structures of regions of main chain as well as side-chains. Therefore, though reassuring about the constancy of the main structural features of proteins, our results emphasize the need for caution in interpreting the biochemical significance of possibly perturbed features of the surface structure.

We are grateful to Professor David Phillips for providing us with data on tetragonal lysozyme and for helpful discussions, to Dr Bob Diamond for providing his real-space refinement programme, and refined tetragonal co-ordinates, and to Drs Gerson Cohen, Leslie Leiserowitz and Michael Levitt for helpful discussions. We would also like to express our appreciation of research grants from the Volkswagen Stiftung (AZ11 11-1790) and the Minerva Foundation and of a fellowship to one of us (J. M.) from the European Molecular Biology Organisation.

REFERENCES

Arndt, U. W. & Willis, B. T. (1966). In *Single Crystal Diffractometry*, pp. 268–276, Cambridge University Press, London.

Beddell, C. R., Moult, J. & Phillips, D. C. (1970). In *Molecular Properties of Drug Receptors* (Porter, R. & O'Connor, M., eds), pp. 85–112, J. and A. Churchill, London.

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.

Blake, C. C. F., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1967*a*). *Proc. Roy. Soc. ser. B*, **167**, 365–377.

Blake, C. C. F., Johnson, L. N., Mair, G. A., North, A. C. T., Phillips, D. C. & Sarma, V. R. (1967*b*). *Proc. Roy. Soc. ser. B*, **167**, 378–388.

Blow, D. M. & Steitz, T. A. (1970). *Annu. Rev. Biochem.* **39**, 63–100.

Browne, W. J., North, A. C. T., Phillips, D. C., Brew, K., Vanaman, C. T. & Hill, R. L. (1969). *J. Mol. Biol.* **42**, 65–86.

Buerger, M. J. (1964). *The Precession Method*, John Wiley, New York.

Coppens, P., Leiserowitz, L. & Rabinovich, D. (1965). *Acta Crystallogr.* **18**, 1035–1038.

Deisenhofer, J. & Steigemann, W. (1975). *Acta Crystallogr. sect. B*, **31**, 238–250.

Diamond, R. (1966). *Acta Crystallogr.* **21**, 253–266.

Diamond, R. (1974). *J. Mol. Biol.* **82**, 371–391.

Drenth, J., Hol, W. G. J., Jansonius, J. N. & Koekoek, R. (1971). *Cold Spring Harbor Symp. Quant. Biol.* **36**, 107–116.

Henderson, R. & Moffat, J. K. (1971). *Acta Crystallogr. sect. B*, **27**, 1414–1420.

Imoto, T., Johnson, L. N., North, A. C. T., Phillips, D. C. & Rupley, J. A. (1972). In *The Enzymes* (Boyer, P. D., ed.), vol. 7, pp. 665–868, American Press, New York.

IUPAC-IUB (1970). *J. Mol. Biol.* **52**, 1–17.

Joynson, M. A., North, A. C. T., Sarma, V. R., Dickerson, R. E. & Steinrauf, L. K. (1970). *J. Mol. Biol.* **50**, 137–142.

Levitt, M. (1974). *J. Mol. Biol.* **82**, 393–420.

Luzzati, V. (1953). *Acta Crystallogr.* **6**, 142–152.

Moult, J., Yonath, A., Rabinovich, D. & Traub, W. (1973). *Stockholm Symp. on Structure of Biol. Molecules*, Abstr. p. 77.

North, A. C. T., Phillips, D. C. & Mathews, F. S. (1968). *Acta Crystallogr. sect. A*, **24**, 351–359.

Phillips, D. C. (1967). *Proc. Nat. Acad. Sci., U.S.A.* **57**, 484–495.

Phillips, D. C. (1971). In *The Harvey Lectures*, vol. 66, pp. 135–160, Academic Press, New York and London.

Rabinovich, D. & Schmidt, G. M. J. (1973). U.S. Patent 3,728.451.

Richards, F. M. (1968). *J. Mol. Biol.* **37**, 225–230.

Rollett, J. S. & Sparks, R. A. (1960). *Acta Crystallogr.* **13**, 273–274.

Rossmann, M. G. & Blow, D. M. (1962). *Acta Crystallogr.* **15**, 24–31.

Srinivasan, R. & Ramachandran, G. N. (1965). *Acta Crystallogr.* **19**, 1008–1014.

Steinrauf, L. K. (1959). *Acta Crystallogr.* **12**, 77–79.

Stout, G. H. & Jensen, L. H. (1968). *X-ray Structure Determination*, pp. 360–362, Macmillan, New York.

Tulinsky, A., Vandlen, R. L., Morimoto, C. N., Venkit Mani, N. & Wright, L. H. (1973). *Biochemistry*, **12**, 4185–4192.

Vandlen, R. L. & Tulinsky, A. (1973). *Biochemistry*, **12**, 4193–4200.

Watenpaugh, K. D., Sieker, L. C., Herriot, J. R. & Jensen, L. H. (1973). *Acta Crystallogr. Sect. B*, **29**, 943–956.

Yonath, A., Smilansky, A., Moult, J. & Traub, W. (1973). *9th Int. Congr. Biochem., (Stockholm)*, Abstr. p. 120.