We thank Drs S. H. Banyard, C. C. F. Blake and I. D. A. Swan for allowing us access to their data. P. E. N. is grateful to the Medical Research Council, London, and to Wolfson College, Oxford, for financial support.

References

BANYARD, S. H., BLAKE, C. C. F. & SWAN, I. D. A. (1974). In *Lysozyme* edited by E. OSSERMAN, R. E. CANFIELD & S. BEYCHOK. New York: Academic Press.

BLAKE, C. C. F., KOENIG, D. F., MAIR, G. A., NORTH, A. C. T., PHILLIPS, D. C. & SARMA, V. R. (1965). *Nature, Lond.* **206**, 757–761.

BLAKE, C. C. F. & SWAN, I. D. A. (1971). *Nature, New Biol.* **232**, 12–15.

HACKERT, M. L., FORD, G. C. & ROSSMANN, M. G. (1973). *J. Mol. Biol.* **78**, 665–673.

HUBER, R. (1970). *Crystallographic Computing*, edited by F. R. AHMED, pp. 96–102. Copenhagen: Munksgaard.

HUBER, R., KUKLA, D., BODE, W., SCHWAGER, P., BARTELS, K., DEISENHOFER, J. & STEIGEMANN, W. (1974). *J. Mol. Biol.* **89**, 73–101.

IMOTO, I., JOHNSON, L. N., NORTH, A. C. T., PHILLIPS, D. C. & RUPLEY, J. A. (1972). *The Enzymes.* Vol. VII, 3rd ed. pp. 665–868. Edited by P. BOYER. New York: Academic Press.

JOYNSON, M. A., NORTH, A. C. T., SARMA, V. R., DICKERSON, R. E. & STEINRAUF, L. K. (1970). *J. Mol. Biol.* **50**, 137–142.

LATTMAN, E. (1972). *Acta Cryst.* B**28**, 1065–1068.

LUZZATI, V. (1952). *Acta Cryst.* **5**, 802–810.

NIXON, P. E. (1976). *Acta Cryst.* A**32**, 344.

NIXON, P. E. & NORTH, A. C. T. (1976). *Acta Cryst.* A**32**, 325–333.

NORDMAN, C. E. (1972). *Acta Cryst.* A**28**, 134–143.

PARTHASARATHY, S. & PARTHASARATHI, V. (1972). *Acta Cryst.* A**28**, 426–432.

ROSSMANN, M. G. & BLOW, D. M. (1962). *Acta Cryst.* **15**, 24–31.

TOLLIN, P. (1966). *Acta Cryst.* **21**, 613–614.

TOLLIN, P. (1969). Private communication.

TOLLIN, P. & ROSSMANN, M. G. (1966). *Acta Cryst.* **21**, 872–876.

VAND, V. & PEPINSKY, R. (1956). *Z. Kristallogr.* **108**, 1–14.

WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.

Acta Cryst. (1976). A**32**, 325

Crystallographic Relationship between Human and Hen-Egg Lysozymes. II. Weighting of Electron-Density Maps Phased from an Incomplete Model Structure and Comparison with Map Obtained by Isomorphous Replacement

BY P. E. NIXON* AND A. C. T. NORTH†

Laboratory of Molecular Biophysics, South Parks Road, Oxford OX1 3PS, England

(*Received* 10 *September* 1975; *accepted* 22 *September* 1975)

Methods suitable for completing and refining a protein structure are investigated both theoretically and with human and hen lysozymes as an example. Sim (or Woolfson) weighting and the α-synthesis are compared with unweighted maps, and while the former is an improvement over unweighted maps, the α-synthesis is less clearly an improvement. Both $WF_o \exp(i\alpha_c)$ and difference maps were found to be useful, and a comparison between the isomorphous replacement map of human lysozyme and our maps based on hen lysozyme was encouraging.

1. Introduction

In the previous paper we proposed a model for the structure of human lysozyme, based on the known molecular structure of hen-egg-white lysozyme; we showed that it was possible to determine the orientation and translation of this model to give a reasonable fit to the measured X-ray diffraction intensities of human lysozyme crystals, both at low (6 Å) and medium (2·5 Å) resolutions. It is more interesting, however, to investigate whether or not we can describe the difference between the lysozymes of two species without further phase information. The human/hen lysozyme situation is particularly useful as a test of methods because the human lysozyme structure was being determined by isomorphous replacement methods simultaneously with the work described in this paper (Banyard, Blake & Swan, 1974).

2. Weighting schemes

We have a set of observed structure factor amplitudes, F_o, and a corresponding set of structure factors $\mathbf{F}_c$ (complex quantities) calculated from only some of the atoms in the crystal. This situation is equivalent to that encountered in the 'heavy-atom' method, for which Woolfson (1956) and Sim (1959, 1960) have proposed weighting schemes for centrosymmetrical

* Present address: Department of Chemistry, University of Auckland, Auckland, New Zealand.

† Present address: Astbury Department of Biophysics, University of Leeds, Leeds LS2 9JT, England.

and non-centrosymmetrical crystals, respectively. They find that the correct structure will be represented with least error (r.m.s. over the whole cell) in maps with Fourier coefficients:

non-centrosymmetrical

$$F_o[I_1(X)/I_0(X)] \exp(i\alpha_c)\,, \tag{1}$$

centrosymmetrical

$$F_o \tanh(X/2) \exp(i\alpha_c)\,, \tag{2}$$

where

$$X = 2F_oF_c/\Sigma_Q\,, \tag{3}$$

$$\Sigma_Q = \sum_{\substack{\text{missing}\\ \text{atoms}}} f_i^2 \tag{4}$$

and I_0 and I_1 are Bessel functions.

2(*a*) *Determination of* Σ_Q

Rossmann & Blow (1961) pointed out that Σ_Q should contain a contribution from the inaccuracies in the supposedly known coordinates, as well as that from missing atoms. Their expression for Σ_Q is difficult to use because it includes estimates of these inaccuracies, which are unknown in our case. Σ_Q may be calculated however from the observed and calculated structure factor amplitudes as follows.

We define $\mathbf{F}_Q$ as the difference between the true structure factor $\mathbf{F}$ and that calculated from the model, $\mathbf{F}_c$

$$\mathbf{F}_Q \equiv \mathbf{F} - \mathbf{F}_c\,.$$

$\mathbf{F}_Q$ is thus the Fourier transform of the difference between the true and model structures; this difference will contain negative regions where the model atoms have been wrongly placed. Nixon (1973) showed that if most of the positional errors of the atoms in the model are small compared to $\lambda/2 \sin\theta$ then we may expect $\mathbf{F}_c$ and $\mathbf{F}_Q$ to be uncorrelated, both in phase and magnitude. In this case Srinivasan (1968) has shown (in another context) that the expectation value of F_Q^2 is given by

$$\langle F_Q^2 \rangle = F_o^2 + F_c^2 - 2F_oF_cI_1(X)/I_0(X)\,, \tag{5}$$

where X is given by (3). We have shown (Nixon & North, 1976) that F_o and F_c for human lysozyme follow Wilson's (1949) distribution functions provided that the value of Σ used in these distributions is given by $\Sigma = \langle F^2 \rangle$. It seems reasonable to assume that F_Q will also be distributed in this way, and so the value of Σ_Q that we require for (3) is better estimated by the mean of F_Q^2 than by (4). Summing (5) over all structure factors in a given shell of reciprocal space,

$$\sum_{\mathbf{h}} F_Q^2 = \sum_{\mathbf{h}} \Sigma_Q = \sum_{\mathbf{h}} [F_o^2 - 2F_oF_cI_1(X)/I_0(X)]\,. \tag{6}$$

Since X is given by (3), we see that (6) contains one unknown, Σ_Q, and may be solved by numerical methods. This gives Σ_Q on the same scale as the F_o.

2(*b*) *Comparison of weighting schemes*

If Σ_Q is large, *i.e.* much of the structure remains unknown, X will always be small, and it may be shown that

$$I_1(X)/I_0(X) \approx \tanh(X/2) = \tfrac{1}{2}X + \ldots$$

when $X < ca.$ 1·0. In this case both Woolfson and Sim weights lead to maps with Fourier coefficients:

$$F_o^2F_c \exp(i\alpha_c)\,. \tag{7}$$

This is what Ramachandran & Raman (1959) call the 'α-synthesis' and Buerger (1959) the 'image-seeking sum function', although their derivations are not that given above.

If however Σ_Q is small, *i.e.* most of the structure is known, X will always be large, and (1) and (2) both reduce to

$$F_o \exp(i\alpha_c)\,,$$

the familiar F_o map.

3. Expected effect on densities at atom sites

The arguments in this section are given for the calculation of maps which represent the whole structure, but apply equally well to maps which represent the difference between the true and model structures, *i.e.* to difference maps with coefficients $(F_o - F_c) \exp(i\alpha_c)$ and to Sim-weighted difference maps with coefficients $\{[F_oI_1(X)/I_0(X)] - F_c\} \exp(i\alpha_c)$.

Luzzati (1953) assumed that $\mathbf{F}_c$ and $\mathbf{F}_Q$ are independent in magnitude and phase, and that their magnitudes are distributed according to Wilson's (1949) acentric distribution. He showed that on a map with Fourier coefficients $F_o \exp i\alpha_c$ the expected electron density at the known atomic sites is τ times the correct electron density, where

$$\tau = \int_0^\infty \int_0^\infty 4\psi^2x^2y^2 \exp[-x^2 - y^2(1+\psi)]I_0(2xy)\mathrm{d}x\mathrm{d}y\,, \tag{8}$$

and where

$$x \equiv F_o\Sigma_Q^{-1/2}$$

$$y \equiv F_c\Sigma_Q^{-1/2}$$

$$\psi \equiv \Sigma_Q/\Sigma_c$$

$$\Sigma_c \equiv \sum_{\substack{\text{known}\\ \text{atoms}}} f_i^2\,.$$

The electron density at the unknown atom sites is χ times the correct electron density, where

$$\chi = \int_0^\infty \int_0^\infty \{4\psi x^2y[xI_1(2xy) - yI_0(2xy)] \times \exp[-x^2 - y^2(1+\psi)]\}\mathrm{d}x\mathrm{d}y\,. \tag{9}$$

Luzzati derived similar expressions for the case of Wilson's (1949) centric distribution. He further pointed

out that while the mean square electron density is $\langle F^2\rangle=\Sigma$, the electron density accounted for by the above (*i.e.* at known and unknown atom sites) is

$$\chi^2 \sum_{\substack{\text{unknown}\\ \text{atoms}}} f_i^2+\tau^2 \sum_{\substack{\text{known}\\ \text{atoms}}} f_i^2=\Sigma_Q(\chi^2+\tau^2/\psi)\,.$$

If the remaining electron density is noise, the signal/r.m.s. noise ratio is

$$\mu\equiv\chi[(\langle F^2\rangle/\Sigma_Q)-(\chi^2+\tau^2/\psi)]^{-1/2}\,. \qquad (10)$$

3(*a*) *Extension to weighted case*

If we wish to calculate χ and τ for a map the amplitudes of whose Fourier coefficients are not simply F_o, we replace F_o by WF_o or $F_o^2F_c$ in Luzzati's (1953) equations (1) and (10) or (24) and (36). The effect of applying Sim weighting is given for example in (8) of this paper by multiplying the integrand by $I_1(2xy)/I_o(2xy)$.

The expressions for χ, τ and $\langle F^2\rangle$ are collected together below, where $E=\exp\{-[x^2+y^2(1+\psi)]/2\}$, $S=\sinh(xy)$, $C=\cosh(xy)$, $I_1=I_1(2xy)$, $I_0=I_0(2xy)$ (Bessel functions), and all integrals are from 0 to ∞.

Unweighted case

$$\tau=\iint 4\psi^2x^2y^2E^2I_0\mathrm{d}x\mathrm{d}y\,, \quad \textit{acentric} \qquad (11)$$

$$\tau=\iint\frac{2}{\pi}\psi^{3/2}xyEC\mathrm{d}x\mathrm{d}y\,, \quad \textit{centric} \qquad (12)$$

$$\chi=\iint 4\psi x^2yE^2(xI_1-yI_0)\mathrm{d}x\mathrm{d}y\,, \quad \textit{acentric} \qquad (13)$$

$$\chi=\iint\frac{2}{\pi}\psi^{1/2}xE(xS-yC)\mathrm{d}x\mathrm{d}y\,, \quad \textit{centric} \qquad (14)$$

$$\langle F^2\rangle=\iint\Sigma_Q4\psi x^3yE^2I_0\mathrm{d}x\mathrm{d}y = \left(1+\frac{1}{\psi}\right)\Sigma_Q\,, \quad \textit{acentric} \qquad (15)$$

$$\langle F^2\rangle=\iint\Sigma_Q\frac{2}{\pi}\psi^{1/2}x^2EC\mathrm{d}x\mathrm{d}y = \left(1+\frac{2}{\psi}\right)\Sigma_Q\,, \quad \textit{centric}\,. \qquad (16)$$

Sim weighting (acentric case)

$$\tau=\iint 4\psi^2x^2y^2E^2I_1\mathrm{d}x\mathrm{d}y=1\,, \qquad (17)$$

$$\chi=\iint 4\psi x^2yE^2(xI_1^2/I_0-yI_1)\mathrm{d}x\mathrm{d}y\,, \qquad (18)$$

$$\langle F^2I_1^2/I_0^2\rangle=\iint\{\Sigma_Q4\psi x^3yE^2I_1^2/I_0\}\mathrm{d}x\mathrm{d}y\,. \qquad (19)$$

Woolfson weighting (centric case)

$$\tau=\iint\frac{2}{\pi}\psi^{3/2}xyES\mathrm{d}x\mathrm{d}y=1\,, \qquad (20)$$

$$\chi=\iint\frac{2}{\pi}\psi^{1/2}xE(xS^2/C-yS)\mathrm{d}x\mathrm{d}y, \qquad (21)$$

$$\langle F_o^2\tanh^2(xy)\rangle=\iint\{\Sigma_Q\frac{2}{\pi}\psi^{1/2}x^2ES^2/C\}\mathrm{d}x\mathrm{d}y\,. \qquad (22)$$

α-synthesis

$$\tau=\iint\Sigma_Q4\psi^2x^3y^3E^2I_0\mathrm{d}x\mathrm{d}y = \left(1+\frac{2}{\psi}\right)\Sigma_Q\,, \quad \textit{acentric} \qquad (23)$$

$$\tau=\iint\Sigma_Q\frac{2}{\pi}\psi^{3/2}x^2y^2EC\mathrm{d}x\mathrm{d}y = \left(1+\frac{3}{\psi}\right)\Sigma_Q\,, \quad \textit{centric} \qquad (24)$$

$$\chi=\iint\Sigma_Q4\psi x^3y^2E^2(xI_1-yI_0)\mathrm{d}x\mathrm{d}y =\Sigma_Q/\psi\,, \quad \textit{acentric} \qquad (25)$$

$$\chi=\iint\Sigma_Q\frac{2}{\pi}x^2yE(xS-yC)\mathrm{d}x\mathrm{d}y =2\Sigma_Q/\psi\,, \quad \textit{centric} \qquad (26)$$

$$\langle F_o^4F_c^2\rangle=\iint\Sigma_Q^34\psi x^5y^3E^2I_0\mathrm{d}x\mathrm{d}y =2\Sigma_Q^3\frac{1}{\psi}\left(1+\frac{1}{\psi}\right)\left(1+\frac{3}{\psi}\right), \quad \textit{acentric} \qquad (27)$$

$$\langle F_o^4F_c^2\rangle=\iint\Sigma_Q^3\frac{2}{\pi}\psi^{1/2}x^4y^2EC\mathrm{d}x\mathrm{d}y =3\Sigma_Q^3\frac{1}{\psi}\left(1+\frac{1}{\psi}\right)\left(1+\frac{5}{\psi}\right), \quad \textit{centric}\,. \qquad (28)$$

The second parts of (23) and (25) were derived from an assumption of point atoms by Raman (1959). From (10) and (23) to (28) we see that the signal-to-noise ratio μ for an α-synthesis is given in the acentric case by

$$\mu=(2/\psi+3+\psi)^{-1/2}$$

and in the centric case by

$$\mu=2(6/\psi+8+2\psi)^{-1/2}\,.$$

3(*b*) *Results*

Equations (11) to (28) were integrated numerically, and those functions for which a simple algebraic expression is not apparent are listed in Table 1.

[A careful comparison of the results for the unweighted, acentric case in Table 1, with those in Table 1 of Luzzati's (1953) paper will reveal a slight discrepancy: re-evaluation of the values for χ and τ by Luzzati's method yields the results of Table 1.] Equa-

tions (17) and (20) show that with suitable weighting the electron density at the known atom sites is correct, and therefore no 'ghost' of the model structure will be seen on a weighted difference map.

Fig. 1 (top) shows the signal-to-noise ratio μ for the unweighted, Sim or Woolfson weighted, and α-synthesis maps, for both the centric and acentric cases. The weighted maps are better than the unweighted by this criterion, though the improvement is perhaps not startling. The α-synthesis is similar to the weighted synthesis when only a small proportion of the structure is known (σ_1^2 less than about 0·2), as is discussed in §2(*b*), but is less good as σ_1^2 increases.

Fig. 1 also shows χ/τ, the ratio of unknown/known atom electron density divided by their atomic numbers. On this criterion the α-synthesis is the best, with the weighted maps still showing an improvement over the unweighted maps. We feel that this criterion is not as useful, however, as the signal-to-noise ratio.

4. Human and hen lysozymes

A set of 858 atoms whose coordinates were derived from the structure of hen-egg-white lysozyme fits the diffraction data from human lysozyme quite well, with an average error on the atomic positions of the order of 0·6 Å (Nixon & North, 1976). This model does not contain any of the atoms that differ between human and hen lysozymes as a consequence of amino acid substitutions, and our primary objective in the present work was to see whether we could derive an electron density map for human lysozyme that would disclose the missing atoms. We also were concerned to see whether our maps would give any indication that amino acid side chains common to the two species were ever in different positions, contrary to the assumption made in our model.

We have used two essentially different types of method in endeavouring to improve our initial 858 atom model. The first type is based on 'direct methods' of phasing and aims at modifying the initial phases in such a way as to make the electron density map conform more nearly to expectation; thus, electron density should nowhere be negative, nor for a protein should there be any particularly high peaks. The methods we have used are based on modification of the electron density map, followed by Fourier transformation to give a new set of phases and the calculation of a new map. They are closely related to the reciprocal-space formulations of Sayre (1952) and of Karle & Karle (1966) (the 'tangent formula').

5. Direct methods refinement

5(*a*) *Lower truncation – the elimination of negative electron density*

Fig. 2 illustrates various types of map modification methods that have been used in attempts to refine

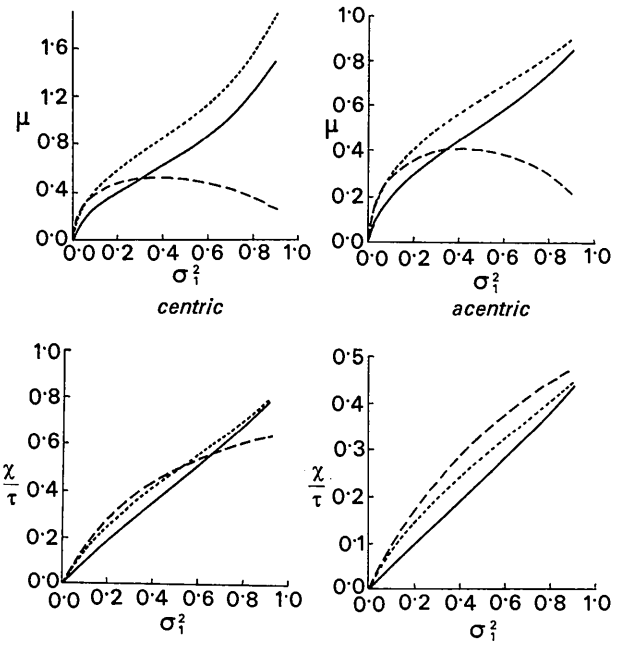

Fig. 1. A comparison of the unweighted (solid line), weighted (dotted line) and α-synthesis (dashed line) maps. μ is the signal-to-noise ratio, and χ/τ is the (unknown atom/known atom) electron density ratio.

Table 1. *Numerical values for the integrals, equations* (13), (11), (18), (19), (14), (12), (21), (22), *as a function of* $\sigma_1^2 = \sum_c/(\sum_c + \sum_Q)$

	Acentric				Centric			
	Unweighted		Sim weighting		Unweighted		Woolfson weighting	
σ_1^2	χ	τ	χ	$\langle W^2F_o^2\rangle/\Sigma$	χ	τ	χ	$\langle W^2F_o^2\rangle/\Sigma$
0	0		0	0	0		0	0
0·1	0·126	2·546	0·081	0·173	0·205	2·115	0·149	0·234
0·2	0·180	1·845	0·143	0·314	0·295	1·568	0·255	0·404
0·3	0·224	1·544	0·195	0·436	0·362	1·341	0·343	0·540
0·4	0·263	1·369	0·241	0·545	0·436	1·216	0·421	0·631
0·5	0·300	1·255	0·283	0·642	0·500	1·137	0·493	0·747
0·6	0·335	1·173	0·324	0·730	0·564	1·084	0·563	0·825
0·7	0·370	1·112	0·363	0·809	0·631	1·048	0·635	0·890
0·8	0·406	1·065	0·403	0·881	0·705	1·023	0·711	0·942
0·9	0·446	1·029	0·445	0·944	0·793	1·005	0·799	0·982
1·0	0·5	1	0·5	1	1	1	1	1

structures. Our first attempt is summarized in Fig. 2(*b*). We took the set of 858 atoms described above as our first approximation to the structure of human lysozyme and used them to calculate phases for an unweighted Fourier synthesis of the 4089 measured structure-factor amplitudes (kindly given us by Drs. Banyard and Blake). We did not know the true zero level of this map [having omitted the $F(0,0,0)$ term], and chose a level of -70 on an arbitrary scale as the zero. Electron density thus designated as negative and comprising 23·5% of the unit cell was then set to zero. We calculated the Fourier transform of this adjusted electron density map with a Fortran subroutine given us by Lifchitz (1974), at grid intervals of $\frac{1}{64}$ in x and y and $\frac{1}{32}$ in z, *i.e.* about 1 Å in each dimension.

The Fourier transform of the modified electron density gave phases that differed on average by 5·9° from the starting set.

At this stage of our work, we were able to compare our phases with those that had recently been obtained by the isomorphous replacement method in order to see whether our attempts to improve them had led to convergence between the two sets. The initial mean phase difference was 69·1°; after this one cycle of refinement the difference was 69·4°. This did not encourage us to try another similar cycle of refinement, but we then tried an upper and a lower truncation, as in Fig. 2(*d*).

5(b) Double truncation – limitation of high electron density

The upper truncation level in electron density was set at $+70$ (the lower cut-off level was maintained at -70) and this further truncation affected 18·6% of the volume of the unit cell. After this double truncation and Fourier transformation the phases were found to have changed by an average of 21·6°. The mean discrepancy with the isomorphous replacement phases was 71·6°: *i.e.* the phase discrepancy had been increased by an average of 2·5°. This again did not encourage us to try another cycle.

5(c) Double truncation and squaring – enhancement of 'signal-to-noise' ratio.

The next modification we tried is illustrated in Fig. 2(*f*). The cut-off levels were the same as for the double truncation just described. The mean phase change was 20·4° and the new phase set differed from the isomorphous replacement phases by 71·4°, an increase of 2·3°. This again does not indicate that the Fourier-modify-Fourier method is improving the phases.

The mean phase differences for these three attempts were also calculated as functions of $\sin\theta/\lambda$. The differences between the isomorphous replacement phases and the phases deriving from our model structure increased with $\sin\theta/\lambda$. This is partly a reflexion of the decrease of figure of merit of the isomorphous replacement phases with $\sin\theta/\lambda$, and is partly because our model seems less accurate at higher resolution, which is to be expected in view of the detailed differences between the two molecules. The only instance of an improvement in our phases was at low resolution ($\lambda/2\sin\theta > 12$ Å) for the double truncation and squaring modification of the electron density. These low-resolution phases showed a mean change of only 8° after truncation and squaring, and the new set differed from the isomorphous replacement set by 2° *less* after this cycle of refinement than before.

We concluded however that this type of refinement was less profitable than would be the study of weighted electron density and difference maps.

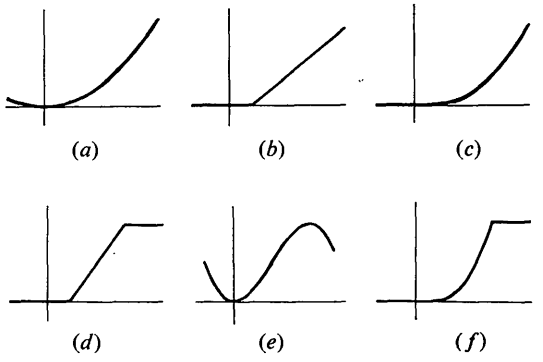

Fig. 2. Some modifications to approximate electron densities which have been used to improve phases by the Fourier transform-modify-Fourier transform method. (*a*) squaring [Sayre, (1952) and many others], (*b*) lower truncation (Kartha, 1969), (*c*) lower truncation plus squaring (Barrett & Zwick, 1971), (*d*) lower and upper truncation (Gassmann & Zechmeister, 1972), (*e*) cubic approximation to (*d*) (Hoppe & Gassmann, 1968), (*f*) lower and upper truncation plus squaring (this work). Only modifications (*b*), (*d*) and (*f*) were used in this work.

6. Hen–human lysozyme difference maps

Four Fourier syntheses were calculated incorporating structure factors $F_c \exp i\alpha_c$ calculated as described in §4. The Fourier coefficients were

(i) $F_o \exp i\alpha_c$.

(ii) $WF_o \exp i\alpha_c$, where W is the Sim or Woolfson weighting factor [see (1) and (2)].

(iii) $F_o^2 F_c \exp i\alpha_c$, the α-synthesis.

(iv) $(WF_o - F_c) \exp i\alpha_c$, a weighted difference map.

For the calculation of Σ_Q for use in the weighting functions, equation (6) was solved for Σ_Q numerically. Four shells of reciprocal space were taken, with limits at 6, 4, 3 and 2·5 Å, and a value of Σ_Q obtained for each. The ratio σ_1^2 (the ratio of known to total scattering matter) was found to be 0·54, 0·55, 0·49 and 0·40 respectively for these shells of data; the decrease of σ_1^2 with increasing resolution is due to the increasing contribution to Σ_Q of errors in the coordinates of the model. Taking σ_1^2 as 0·5 we see from Fig. 1 that

(*a*) For map (i), the signal-to-noise ratio μ will be 0·52 and the unknown/known electron density χ/τ will be 0·24.

(*b*) For map (ii), $\mu=0{\cdot}63$ and $\chi/\tau=0{\cdot}28$, which should be a noticeable improvement over map (i).

(*c*) For map (iii), $\mu=0{\cdot}41$ and $\chi/\tau=0{\cdot}33$.

(*d*) For map (iv) it was shown in §3 that there should be no 'ghost' of the model on this weighted difference map; $\mu=0{\cdot}63$ as for map (ii).

6(*a*) *Comparison with isomorphous replacement map*

Fig. 3 compares the section $z=\frac{26}{40}$ through the molecule for map (ii) and the isomorphous replacement map kindly made available to us by Banyard and Blake (the bottom of the active site cleft is on the left of the diagrams). These two maps show a considerable similarity: the main differences are that the α_c map has fewer features in the intermolecular liquid region (top left and bottom right of diagrams) and that the isomorphous replacement map has a feature to the right of centre not found on the α_c map.

Fig. 3 also shows the atoms from which the phases were calculated. The electron density follows these atoms fairly closely but not exactly: for example, there is very little electron density associated with the side chain of residue lysine 1 (top centre), which has now been shown by the isomorphous replacement map to have a quite different orientation in human lysozyme.

Our model had two or more atoms missing from the side chains of 18 residues. In 9 of these a positive region on map (iv) could be seen in a similar position to the side chain on the isomorphous replacement map. Those side chains that appeared less distinct on the isomorphous replacement map – such as tyrosine 62 – were usually not clear on the α_c maps either. There were exceptions to this rule however: the spur to the upper right of $C_\alpha(4)$ in Fig. 4 is in the appropriate place for an L-amino acid side chain, whereas the side chain of glutamate 4 is not clear on the isomorphous replacement map. This may be because Glu 4 is a heavy-atom site for the isomorphous replacement study, but, as we shall discuss below, our model contained no side-chain atoms for residue 4, which is glycine in hen-egg lysozyme. On the other hand, in our maps the guanidinium group of arginine 41 could be represented by any of three isolated blobs of electron density, whereas a low-level but continuous rod of electron density formed the arginine 41 side chain on the isomorphous replacement map. The feature present to the right of centre of Fig. 3 is the 'new' S atom of methionine 29. As may be seen, no corresponding atom appears on map (ii) (nor on map (iv)).

Side chains usually have a higher temperature factor than the main chain (Watenpaugh, Sieker, Herriott & Jensen, 1973), so that on a difference map calculated from atoms with equal temperature factors, the side chains are likely to appear in holes. The carboxylate group of the side chain of aspartic acid 18, for example, appears in a neat hole on map (iv); map (ii) does not indicate that it has moved, for the atoms lie at the centre of the peak of electron density on this map. When this group was omitted and the phases recalcu-

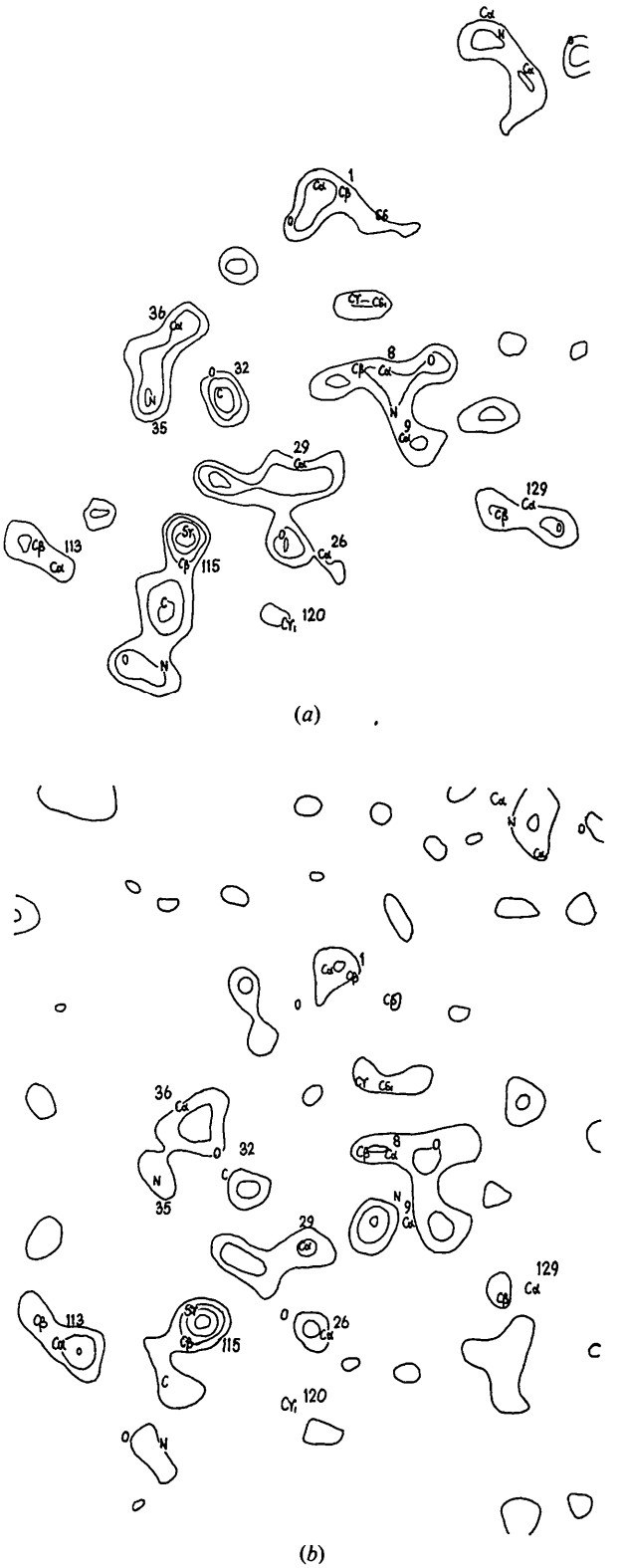

Fig. 3. Section $z=\frac{26}{40}$ of the Sim weighted map phased from 858 atoms (*a*), compared to the isomorphous replacement map (*b*) for human lysozyme. The same set of structure factor amplitudes was used in each case.

lated, electron density reappeared exactly where the carboxylate group had been. This comparison of the difference map with the $WF_o \exp i\alpha_c$ map shows that a higher temperature factor, rather than a change of position, is required.

There are two, and only two, places where a collision occurs between different model molecules when placed in position in the human lysozyme unit cell. They involve the guanidinium groups of arginine residues 14 and 61, which collide with the main chain of residue 126 and with the side chains of valine 109 and asparagine 113 respectively. These are the locations of the two deepest holes in the difference map.

We now discuss four regions in detail.

6(*b*) *Residues* 4, 5 *and* 6

In hen lysozyme these are –Gly–Arg–Cys– and in human they are –Glu–Arg–Cys–; all the hen atoms were used, therefore, in the model for human lysozyme. We should look for a possible movement of these atoms, and for the appearance of the side chain of Glu 4.

Fig. 4 shows the sections containing most of these atoms. Maps (i), (ii) and (iii) look quite similar, with the top of the rod of electron density corresponding to $C_\alpha(4)$ to $C_\alpha(5)$ slightly to the right of the atomic positions used for the phasing. Map (ii) appears to be contoured at a slightly lower level than map (i) because the weights of the Fourier coefficients are always less than unity. Map (iii) is at a lower effective resolution than map (i) because the weighting function F_oF_c tends to decrease with increasing $\sin\theta$. The isomorphous replacement map shows that the model atoms are approximately correct, but that they should be moved to the right by about 0·7 Å, further than a simple interpretation of maps (i) to (iii) would indicate (about 0·3 Å).

Map (iv), the difference map, shows atoms $C_\alpha(4)$ to $C_\beta(5)$ to be to the right of a hole and (less continuously) to the left of a peak. The magnitude of the shift may be given, if small, by (see *e.g.* Stout & Jensen, 1968)

$$\Delta x = -\frac{\partial \Delta\varrho/\partial x}{\partial^2\varrho/\partial x^2} .$$

The curvature was estimated from sections through atoms in the main chain on map (ii), the sections being taken roughly perpendicular to the chain directions (curvatures obtained in this way varied from atom to atom with a standard deviation of about 10%). Stout & Jensen point out that the slope $\partial\varrho/\partial x$ from such a difference map should be divided by 0·5 because the wanted features appear with only half weight in the limit of the model being only very slightly wrong. § 3 and Table 1 of this paper show that we should divide the curvatures by 0·28 (χ when $\sigma_1^2 = 0{\cdot}5$). When this is done, the average shift of the atoms $C_\alpha(4)$ to $C_\beta(5)$ predicted from map (iv) is $0{\cdot}7 \pm 0{\cdot}15$ Å, in agreement with the isomorphous replacement map.

It is not clear from Fig. 4 whether map (ii) is better than map (i) or not.

Fig. 5 shows the bottom contour level of the section $z = \frac{20}{40}$ superimposed for maps (i) and (ii) and for maps (iii) and (ii). The electron density is further to the right on map (ii) than on map (i), indicating that map (ii) is

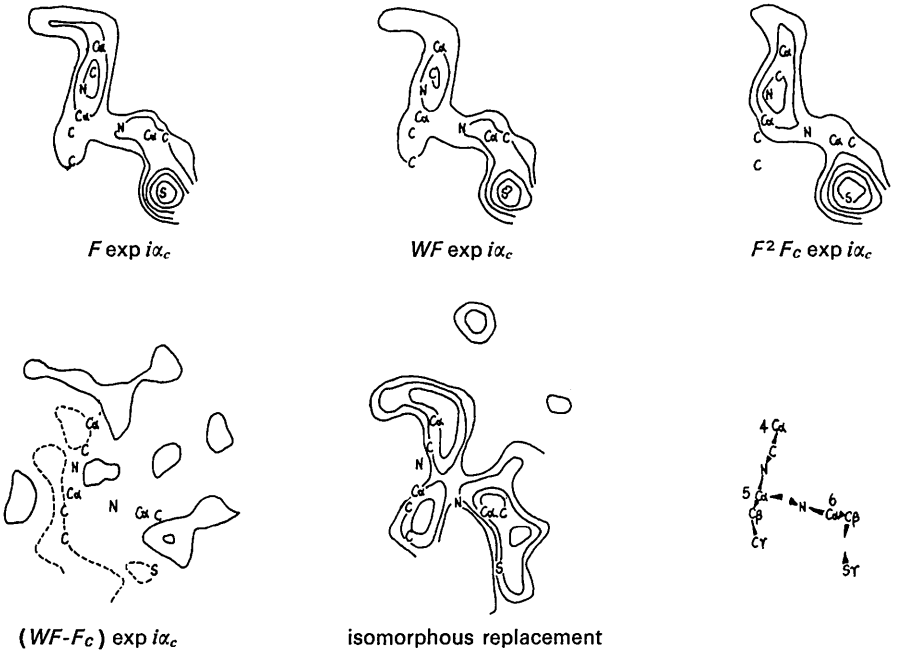

Fig. 4. A superposition of features from the sections $z = \frac{20}{40}$ and $\frac{21}{40}$ for various maps near residues 4, 5 and 6. The geometry of the atoms concerned is indicated bottom right.

better than map (i) in this region. There seems to be no systematic difference between maps (ii) and (iii), as shown in Fig. 5, except for the lower resolution of the latter.

6(*c*) *Residues* 72, 73, 74, 75 *and* 77

Most of the main chain atoms of this surface region of the molecule lie close to the sections $z=\frac{12}{40}$ and $\frac{13}{40}$, and are shown in Fig. 6. Map (iv) shows that many of these atoms lie near the bottom of holes in the difference map, and it is not clear to which positive region these atoms should be moved – especially residues 74 and 75. An examination of map (ii), or better a comparison of maps (ii) and (iv), shows that residues 72 and 73 move down, 77 is approximately correctly placed, and residues 74 and 75 are probably also near their correct positions: their appearance in holes in map (iv) is probably due to their high temperature factor, for this surface region of the molecule has few hydrogen bonds to stabilize its conformation. These conclusions are verified by the corresponding portion of the isomorphous replacement map.

6(*d*) *Residues* 17, 23, 28, 105 *and* 111

These five residues form a hydrophobic region in the interior of the molecule; unfortunately it is difficult to represent in two dimensions. The isomorphous replacement study of Banyard & Blake has shown that these side chains are arranged differently in human and hen lysozymes, and it would be very difficult to predict this rearrangement from the amino-acid sequence only. If our study is to be of value in the determination of unknown protein structures, it must be able to detect such a rearrangement.

According to the isomorphous replacement map the two central residues of this cluster, 23 and 105, become smaller on going from hen to human – changing from tyrosine and methionine to isoleucine. Tryptophan 28 twists inwards, partially to fill this hole, and 17 is enlarged from leucine to methionine, the S atom occupying part of the space occupied by tryptophan 28 in hen lysozyme. On the other side of this hydrophobic core the six-membered ring of tryptophan 111 moves inwards while the bottom of the five-membered ring hinges outward slightly.

When one knows what one is looking for, features in map (iv) can be found to explain the differences described in the last paragraph. The six-membered ring of tryptophan 28 lies in a hole in map (iv), with a peak between its hen position and methionine 17;

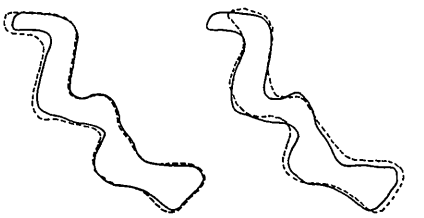

Fig. 5. Comparison of the lowest contour level of section $z=\frac{20}{40}$ in the region of residues 4, 5 and 6. (*a*) $F\exp i\alpha_c$ (dotted) and $WF\exp i\alpha_c$ (solid); (*b*) $F^2F_c\exp i\alpha_c$ (dotted) and $WF\exp i\alpha_c$ (solid).

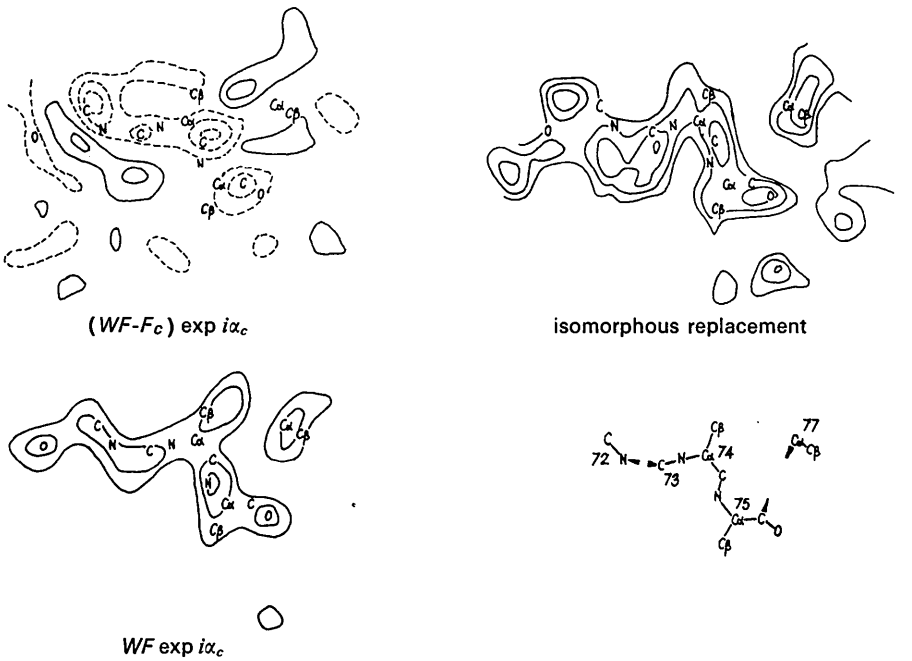

Fig. 6. A superposition of features from the sections $z=\frac{12}{40}$ and $\frac{13}{40}$ for various maps near the residues 72–75 and 77. The geometry of the atoms concerned is indicated bottom right.

this peak represents the new S atom of methionine 17, rather than a movement of tryptophan 28. Although there is no other possible peak for the S of methionine 17, this is not strong enough evidence to force us to the correct interpretation of the difference map. Map (ii) seems to indicate that no change is necessary in this region. The other tryptophan residue which has moved, 111, also lies in a hole in map (iv), with a peak near the five-membered ring which correctly indicates its changed position. There is a peak which represents the new position of the six-membered ring, but since this is about 5 Å away, this positive feature would not necessarily be associated with the negative feature representing the vacated site of the six-membered ring. Map (ii) shows clearly that tryptophan 111 is in a different orientation in human lysozyme, backing up the conclusions from the difference map. In short, the whole region is definitely disturbed, but the new positions of the side chains are not clear from these maps.

7. Conclusions

§ 3 of this paper indicated that the use of the Sim weighting scheme should result in a definite but not outstanding improvement in the maps, and Fig. 5 and the pertinent discussion seem to confirm this. Other similar examples could have been given.

Watenpaugh *et al.* (1973) have shown that it is possible and praticable to refine protein structures; our work differs from theirs in that our starting point is much less accurate and less complete than theirs, and that our structure factors extend only to 2·5 Å resolution, rather than the 1·5 Å available for rubredoxin. We feel that our large shifts and low resolution explain why the atomic shifts were often not clear from the difference map alone. However, the comparison of the $WF_o \exp i\alpha_c$ and difference maps usually gave a clear indication of the atomic shifts in agreement with the isomorphous replacement map. Many of the missing side chains could be placed, and this leads us to believe that to refine our structure completely would present no insuperable problem.

We chose from our original 858 atoms a set of 327 which we felt were almost certainly near their correct positions. Difference maps phased from these 327 atoms showed features at the sites of most of the omitted atoms – indicating that we had been overcautious in omitting them – but with a higher superimposed noise. In the region of the hydrophobic core rearrangement discussed in the last section, this map showed side-chain features in the same places as the isomorphous replacement map, but because of the lower signal-to-noise ratio ($\mu=0{\cdot}49$, as opposed to $\mu=0{\cdot}63$ for the maps phased from 858 atoms) they were sometimes poorly connected to the main chain.

We did not continue the refinement because the result would merely be a duplication of the isomorphous replacement study, but we are confident that it could be done. On the other hand, our attempts to refine the phases by automatic methods related to 'direct methods' of phasing were not successful in this fairly simple application.

We are indebted to Drs. S. H. Banyard, C. C. F. Blake and I. D. A. Swan for access to their measured structure factors and isomorphous replacement phases. P. E. N. is grateful to the Medical Research Council, London, for a scholarship in training in research methods, and to Wolfson College, Oxford, for financial support.

References

BANYARD, S. H., BLAKE, C. C. F. & SWAN, I. D. A. (1974). In *Lysozyme*, edited by E. OSSERMAN, R. E. CANFIELD & S. BEYCHOK. New York: Academic Press

BARRETT, A. N. & ZWICK, M. (1971). *Acta Cryst.* A**27**, 6–11.

BUERGER, M. J. (1959). *Vector Space.* New York: John Wiley.

GASSMANN, J. & ZECHMEISTER, K. (1972). *Acta Cryst.* A**28**, 270–280.

HOPPE, W. & GASSMANN, J. (1968). *Acta Cryst.* B**24**, 97–108.

KARLE, J. & KARLE, I. L. (1966). *Acta Cryst.* **21**, 849–859.

KARTHA, G. (1969). Private communication to BARRETT & ZWICK (1971).

LIFCHITZ, A. (1974). *Acta Cryst.* A**30**, 86–92.

LUZZATI, V. (1953). *Acta Cryst.* **6**, 142–152.

NIXON, P. E. (1973). *A Crystallographic Comparison of Protein Molecules with Related Conformations.* D. Phil. Thesis, Univ. of Oxford.

NIXON, P. E. & NORTH, A. C. T. (1976). *Acta Cryst.* A**32**, 320–325.

RAMACHANDRAN, G. N. & RAMAN, S. (1959). *Acta Cryst.* **12**, 957–964.

RAMAN, S. (1959). *Acta Cryst.* **12**, 964–975.

ROSSMANN, M. G. & BLOW, D. M. (1961). *Acta Cryst.* **14**, 641–647.

SAYRE, D. (1952). *Acta Cryst.* **5**, 60–65.

SIM, G. A. (1959). *Acta Cryst.* **12**, 813–815.

SIM, G. A. (1960). *Acta Cryst.* **13**, 511–512.

SRINIVASAN, R. (1968). *Z. Kristallogr.* **126**, 175–181.

STOUT, G. H. & JENSEN, L. H. (1968). *X-ray Structure Determination: a Practical Guide.* New York: Macmillan.

WATENPAUGH, K. D., SIEKER, L. C., HERRIOTT, J. R. & JENSEN, L. H. (1973). *Acta Cryst.* B**29**, 943–956.

WILSON, A. J. C. (1949). *Acta Cryst.* **2**, 318–321.

WOOLFSON, M. M. (1956). *Acta Cryst.* **9**, 804–810.