Acta Cryst. (1981). A**37**, 491–500

The Use of Phase Combination in the Refinement of Phosphoglycerate Kinase at 2·5 Å Resolution

BY D. W. RICE*

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

(*Received* 16 *June* 1980; *accepted* 14 *January* 1981)

Abstract

The combination of phase information obtained by the method of isomorphous replacement with that calculated from a partial model of the structure has proved to be a powerful tool in the interpretation of the structure of horse muscle phosphoglycerate kinase. That the combined phases are an improvement on the calculated or isomorphous ones is evidenced by more readily interpretable electron density maps, which show only a low level of feedback from incorrectly located atoms. Several different types of Fourier syntheses have been calculated using the combined phase angles and an essentially qualitative evaluation of their relative merits is presented.

Introduction

The starting models of most protein structures are obtained by the inspection of a 3·0 or 2·5 Å resolution electron density map whose phases are calculated by the method of multiple isomorphous replacement (Blow & Rossmann, 1961, and the references therein). These phases are subject to errors which arise from the combination of a number of factors which may include non-isomorphism of the derivatives, low occupancy of the heavy atoms, poor heavy-atom refinement and errors in the measurements of the X-ray intensity data. The normal analysis of these errors leads to the assignment of a figure of merit (Blow & Crick, 1959) to each reflection, this factor expressing the precision of the phase determination. The result of the phase-angle errors and the figure of merit weighting is to degrade the quality of the resultant electron density image with noise and to lower its apparent resolution.

To define the protein structure it is necessary to determine the orientations of the peptide planes. This is a straightforward procedure when the map is good enough to show the 'bulges' in the electron density due to the peptide carbonyl groups. However, the degradation of the Fourier map by errors in the phase angles often masks these features. Moreover, the electron density may be severely disturbed in the vicinity of the heavy-atom binding sites. In the absence of an amino-acid sequence this can then lead to errors in the chain connectivity. The areas where the greatest problems are encountered in interpreting a protein map are those where the protein atoms have high temperature factors or suffer from static disorder owing to the existence of different conformational states. Such atoms are naturally associated with low peak heights in the electron density maps and the phase angle errors further distort the map. It is not surprising that interpretations in these regions are often incorrect.

Clearly, then, to describe the molecular structure at the level of defining all the side-chain positions, an electron density map with much improved definition is required. Various techniques of refinement and phase extension have been used on proteins where X-ray intensities have been available at or near atomic resolution (Hoppe & Gassmann, 1968; Agarwal & Isaacs, 1977; Sayre, 1974; Hendrickson, 1975; de Rango, Mauguen & Tsoucaris, 1975). However, these techniques have not been used to compute new electron density maps at the limited resolution (2·5 Å) available for the work on PGK, and hence their value in this context is as yet undetermined. (The abbreviations used in this paper are given in Table 1.)

As an alternative approach, the procedure of phase combination, suggested originally by Rossmann & Blow (1961) and developed by Hendrickson & Lattman (1970), allows the phases to be improved by the incorporation of independent phase information from such varied sources as direct methods, non-crystallographic symmetry or a partial structure. Hendrickson & Karle (1973) have improved the isomorphous phases for carp muscle calcium-binding protein by the use of

Table 1. *List of abbreviations used*

PGK	Phosphoglycerate kinase
SIR	Single isomorphous replacement
MIR	Multiple isomorphous replacement
EMP	Ethyl mercury phosphate
EMC	Ethyl mercury chloride
LSQ	Least squares

* Present address: Department of Biochemistry, University of Sheffield, Sheffield, England.

0567-7394/81/040491-10$01.00

the tangent formula of Karle & Hauptman (1956) to refine and extend a starting set of 2 Å isomorphous phases to 1·8 Å. Subsequently the tangent formula phase information was combined with the isomorphous phases to compute a new set of phases. Bricogne (1976) has developed a set of programs for the combination of phase information from isomorphous replacement and non-crystallographic symmetry and this technique has been applied very successfully in several cases (Bloomer, Champness, Bricogne, Staden & Klug, 1978; Harrison, Olson, Schutt & Winkler, 1978). The present paper is concerned with the map improvement that can be achieved by the combination of partial structure phase information with that obtained by multiple isomorphous replacement.

The procedure of phase combination

Rossmann & Blow (1961) proposed that the combination of phase information from a partial structure and from isomorphous replacement should be achieved by multiplication of the corresponding phase probability densities. Thus

$$P(\alpha) = P_{\text{iso}}(\alpha)\, P_{\text{par}}(\alpha), \tag{1}$$

where $P(\alpha)$ is the combined probability function and $P_{\text{iso}}(\alpha)$, $P_{\text{par}}(\alpha)$ are the isomorphous and partial structure phase probabilities respectively. A convenient computational method to achieve the combination was suggested by Hendrickson & Lattman (1970) who defined a general probability function of the form

$$P_{ABCD}(\alpha) = \exp\,(N + A\cos\alpha + B\sin\alpha + C\cos 2\alpha + D\sin 2\alpha), \tag{2}$$

where N is a normalization factor.

With this approach the phase probability information can be stored by saving the values of the four numbers A, B, C and D which are called the four phase coefficients. The problem of phase combination then amounts to one of expressing a given type of probability distribution (*i.e.* isomorphous or partial structure) in terms of the function P_{ABCD} followed by multiplication of these probabilities. The latter involves only the addition of the homologous coefficients.

Hendrickson (1971) described a least-squares fitting procedure for recasting the phase probability distribution computed by the method of Blow & Crick (1959) in terms of the phase coefficients *ABCD*. The phase information from a partial structure has been subjected to a probability analysis by Sim (1959). In his formulation Sim showed that the partial phase function $P_{\text{par}}(\alpha)$ for each Fourier coefficient was

$$P_{\text{par}}(\alpha) = K\exp[X\cos(\alpha - \alpha_{\text{calc}})], \tag{3}$$

where α_{calc} is the calculated phase angle from the model,

$$X = 2F_{\text{obs}}F_{\text{calc}}/\beta$$

and

$$K = 1/2\pi I_0(X)$$

where I_0 is the modified zero-order Bessel function.

Sim then suggested a suitable weighting system that was proportional to $P_{\text{par}}(\alpha)$. Following Blow & Crick (1959), Sim (1960) modified his weighting scheme in order to compute the map with the least mean square error. These weights, W, were therefore defined by

$$W = \int_{-\pi}^{\pi} \cos\alpha\, P_{\text{par}}(\alpha)\, d\alpha, \tag{4}$$

$$= \frac{I_1(X)}{I_0(X)}$$

where I_0 and I_1 are respectively the zero-order and first-order Bessel functions. In Sim's discussion, β was evaluated as $\sum f_j^2$ when the summation is for those atoms not included in the phase determination. Rossmann & Blow (1961) suggested an improvement in the estimate of β. In addition to the contribution to β from the unknown atoms, a further contribution was added based on the standard error in position of the known atoms. Thus

$$\beta = \sum_{j=1}^{J} f_j^2 + \tfrac{4}{3}\pi\mathbf{s}^2 \sum_{h=1}^{H} \sigma_2^2 f_h^2, \tag{5}$$

where $\mathbf{s}$ is the reciprocal-lattice vector and σ is the standard error of the hth atom.

An alternative empirical estimate of β was used by Bricogne (1976), where values were obtained from the lack of closure between I_{obs} and I_{calc} in ranges of $\sin^2\theta/\lambda^2$. Thus

$$\beta = \frac{\sum |I_{\text{obs}} - I_{\text{calc}}|}{N_{hkl}}, \tag{6}$$

where N_{hkl} is the number of reflections included in the summation. The latter was chosen as the weighting scheme in the work described here.

This approach to phase combination has been fully described by Bricogne (1976) and the computer program used to evaluate the combined phase probabilities was a modified version of Dr Bricogne's program *COMBINE*.

The use of phase combination on PGK

PGK is an enzyme of molecular weight 45 000 Daltons which crystallizes in the space group $P2_1$ with one molecule in the asymmetric unit and cell dimensions $a = 50{\cdot}8$, $b = 106{\cdot}9$, $c = 36{\cdot}3$ Å, $\beta = 98{\cdot}6°$.

The single polypeptide chain contains 416 amino acids. The initial interpretation of the structure was obtained from a 3·0 Å SIR map (Blake & Evans, 1974). The quality of this electron density map was

such that when a main-chain trace of the enzyme was constructed without the aid of an amino-acid sequence several errors in chain connectivity were made (Banks *et al.*, 1979). Moreover, some of the surface residues were not recognized in the map as belonging to the protein. Features such as 'carbonyl bulges' were not present and the definition of the side chains was unclear.

A subsequent 2·5 Å MIR map (Rice, 1979) represented some improvement in the electron density, though until the amino-acid sequence was available there was still some doubt as to the accuracy of the proposed chain connectivity. In addition, the interpretations in the loops of chain lying between the regions of regular secondary structure have been shown to be grossly in error, with the deletion and insertion of a number of residues being necessary to complete the chain register (Rice, 1979). Furthermore, the precise location of residues was not possible on this map as 'carbonyl bulges' were not evident and the side-chain definition was still poor.

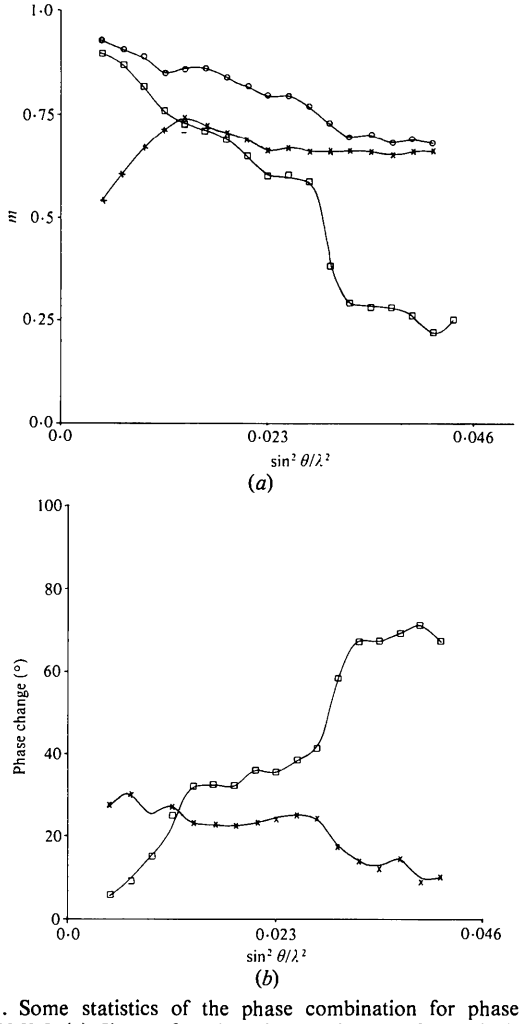


Fig. 1. Some statistics of the phase combination for phase set COMB5. (*a*) Figure of merit: □ isomorphous set 2; × calculated; ○ COMB5. (*b*) Phase change: □ from isomorphous to combined phase; × from calculated to combined phase.

The errors in the phase angles that degraded the quality of the map were due largely to two factors.

(1) The derivatives were all obtained by soaking the crystals in solutions of EMP or EMC, the mercury atoms in these compounds reacting with up to six of the free sulphydryl groups on the enzyme. The reaction conditions were such that the heavy atoms bound at the same sites but with different occupancy patterns. This gave rise to problems in refining the derivatives and also in the fact that the derivatives were not truly independent.

(2) The heavy-metal occupancies were low in some crystals of the derivatives giving rise, therefore, to intensity differences that were comparable with the errors in the measurements. This particularly affected the data in the 3·0–2·5 Å shell [see the isomorphous figure of merit in Fig. 1 (Rice, 1979)].

In an attempt to improve the electron density map the procedure of phase combination was used to combine the phase information available from a progressively improving interpretation of the molecular structure with the isomorphous phase probabilities. The calculated phase information was obtained by placing atoms in the electron density, subjecting them to a refinement process and then calculating their final contributions to the structure factors. Two refinement techniques were used: at first atom shifts were calculated using the technique of automated difference Fourier shifts (see Cochran, 1951) followed by cycles of regularization with the method of Dodson, Isaacs & Rollett (1976). More recently, the Hendrickson–Konnert restrained least-squares refinement program (Konnert, 1976) was employed. Finally, before the phase combination was carried out, group temperature factors (one for each main-chain residue and one for each side chain) were refined by the difference Fourier shift method. This was an attempt to reduce the contribution of the wrongly placed atoms to the calculated phases and hence reduce the feedback. During the work on the crystallographic refinement the amino-acid sequence of the enzyme was determined (see Banks *et al.*, 1979) in the form of 14 sequenced peptides. These were aligned on the X-ray map as they became available and when sufficient criteria for fitting them could be found. The side-chain coordinates of these residues were then introduced into the model.

The phase information from the partial structure has been extracted so far at four stages of the molecular interpretation, and the statistics for phase combination are shown in Fig. 1 and Table 2. Table 3 summarizes the deviation from ideality of the bond distances and bond angles of the model PGK7. Three different types of Fourier syntheses with the combined phases were examined. Their relative merits will now be discussed.

Table 2. *Phase combination on PGK*

$$R^* = \sum_{hkl}^{all} |F_{obs} - F_{calc}| / \sum |F_{obs}|.$$

Phase set	Coordinate set	Number of atoms	Calculated phase set	R^*	Figure of merit	Comments
Isomorphous 1	–	–		–	0·46	2·5 Å MIR phases
COMB1	PGK4	1969	CALC1	0·56	0·54	Initial main-chain trace used as model (poly, ala/gly)
COMB2	PGK5	2472	CALC2	0·45	0·62	Model after difference-Fourier refinement. Approximately half the side chains incorporated.
COMB3	PGK6	3016	CALC3	0·32	0·73	Model after LSQ refinement. Almost all sequence incorporated.
COMB4	PGK7	3047	CALC4	0·30	0·74	Model after some further rebuilding and LSQ refinement.
Isomorphous 2	–	–		–	0·53	Isomorphous phases after phase refinement of the derivatives using phase set COMB4.
COMB5	PGK7	3047	CALC4	0·30	0·76	Phase combination of model at stage COMB4 and Isomorphous set 2.
COMB6	PGK8	2617	CALC5	0·33	0·74	Wrongly placed atoms omitted from model.

Table 3. *Deviation from ideality of the model PGK7*

	Number	r.m.s. Δ (Å)
Bond distances	3093	0·012
Angle distances	4153	0·086
Planar 1–4 distances	971	0·076
Deviation from planes	525	0·037

(1) *The normal combined synthesis*

This was a Fourier synthesis calculated with coefficients $|F_{obs}| m \exp(i\alpha_{combined})$, where m is the figure of merit weight. Subjectively these combined-phase Fourier maps appeared to have greater clarity, less noise and higher resolution than the equivalent isomorphous map. There are then two questions that must be answered. Firstly, to what extent does the calculated structure feedback into the electron density map in the places where the model is incorrect? Secondly, does the procedure produce an improvement in those areas of weak electron density for which the initial interpretation was ambiguous?

In order to answer the question about feedback two different criteria were used. Firstly, some pieces of structure of known position (from the 2·5 Å isomorphous map) were left in the wrong orientation with either their side chain or main chain or both lying outside the electron density envelope. The combined maps were then closely scrutinized in these regions for signs of feedback from the incorrectly placed atoms. An example of this type of check is found in Fig. 2. This shows the electron density in map COMB2 around Tyr 323, whose position was known from the isomorphous map. The electron density feature visible in the combined map correlates exactly with the isomorphous map and there is no electron density near the incorrectly placed phenol ring.

A second type of bias check can be made where the interpretation of the isomorphous electron density is ambiguous. If the model is grossly in error, then the refinement procedure will do little to bring about an improvement in the structure, and the incorrectly placed atoms can become distorted into sterically unreasonable conformations with excessively short van

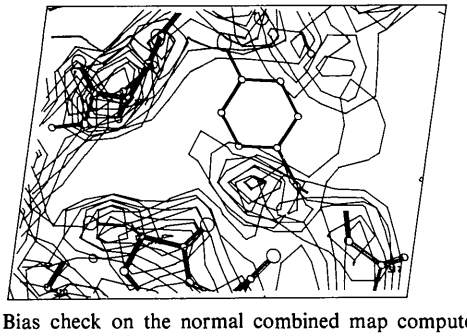


Fig. 2. Bias check on the normal combined map computed with COMB2 phases.

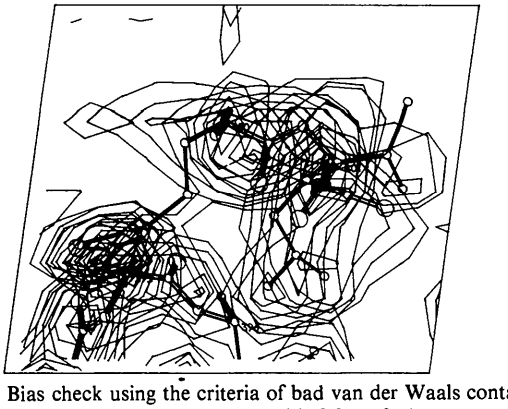


Fig. 3. Bias check using the criteria of bad van der Waals contacts on the normal combined map with COMB3 phases.

der Waals contacts. If there were indeed feedback from the incorrect part of the model, then the observed electron density would reproduce the incorrect conformations. Fig. 3 shows the electron density in map COMB3 around Lys 405. The amino group of the lysine side chain has very bad van der Waals contacts with the carboxyl group of a nearby glutamic acid residue. However, the combined map gives a completely unambiguous indication of the correct position for the lysine side chain.

The regions of weak electron density in the isomorphous map provide excellent checks on the bias and give some estimate of the considerable improvement in the electron density that can be achieved. In an early interpretation of the isomorphous map the side chain of Leu 313 was positioned in the isolated density shown in Fig. 4(*a*). At this time no sequence information was available for residues 346–347 and these were built as Ala Gly.

When the complete sequence became known (Banks *et al.*, 1979), the amino-acid sequence for this pair of residues was found to be Ala Phe. The normal combined synthesis with phase set COMB3 (Fig. 4*b*) revealed two connections in the once isolated density feature around the leucine side chain. At this point it seemed likely that the phenylalanine ring ought to occupy this density feature and the residues 346–347 should be positioned more towards the bottom of the diagram. This would indicate that the electron density at the β-carbon atom of the leucine side chain was due to feedback. In order to see if further improvements in the electron density could be achieved in this region without rebuilding, no alterations to the structure were made for these residues. Instead, other less ambiguous errors in the model were corrected. The normal combined synthesis with phase set COMB5 is illustrated in Fig. 4(*c*). An evident error in the interpretation is now indicated. It is interesting to see that the feedback from the wrongly placed atoms has decreased despite the fact that their contribution to the combined phases has increased owing to the lower R value of the model and hence higher weight of the calculated phase in the phase combination. Fig. 4(*d*) shows a tentative reinterpretation of this part of the structure. Note the fact that the chain has been stepped back one residue. The electron density in Fig. 4(*d*) is the (2–1) synthesis (see next section) with the equivalent phases that were used to calculate Fig. 4(*c*). Thus, the improvement in the electron density that is shown by this figure is due to the different synthesis as the atoms shown in Fig. 4(*d*) were not those that were used in the phase combination.

The indications then from this process of gradual refinement and phase combination are that large initial errors in the interpretation of the structure can be tolerated. In fact, in the model PGK7 some 30% of all the residues can be seen to be significantly misbuilt.

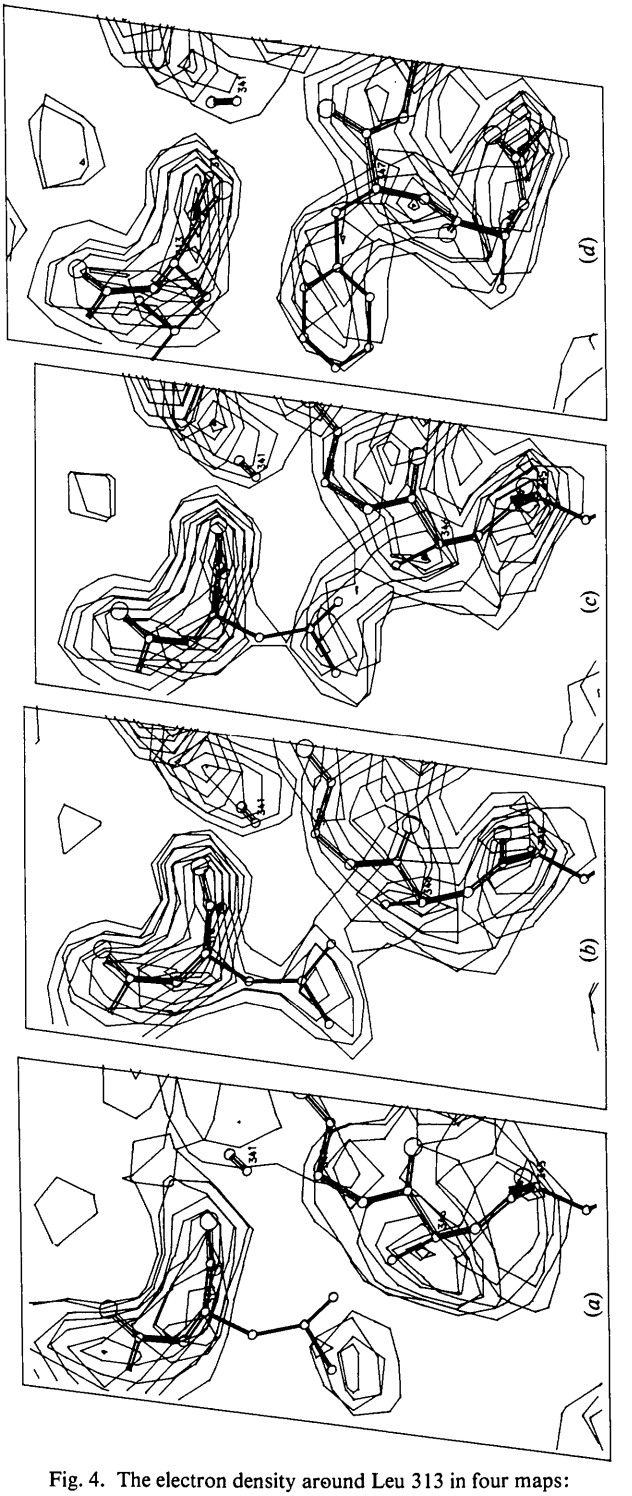


Fig. 4. The electron density around Leu 313 in four maps:

phase set	synthesis type	atoms shown
(*a*) isomorphous 1	–	PGK7
(*b*) COMB3	normal	PGK6
(*c*) COMB5	normal	PGK7
(*d*) COMB5	2–1	tentative reinterpretation.

Whilst some of the errors may merely involve a small rotation of a side chain or main chain of a residue to achieve a better fit to the electron density, some 10% of all the residues are grossly wrong and of the type illustrated by Fig. 4.

The combined maps show much greater main-chain and side-chain definition. Carbonyl bulges are very obvious so that rebuilding is greatly facilitated. The electron density peak heights at the sulphur atoms of methionine and cysteine residues are much stronger than at the positions of the other protein atoms. This was not the case in the isomorphous map. The aromatic rings can be seen in these maps as flat discs rather than the spheroidal shape they had in the isomorphous map, and at the ends of some of the well ordered arginine side chains a bifurcation of the electron density feature, owing to the two NH groups, can be seen. The disturbances in the electron density in the vicinity of the heavy-metal binding sites are greatly reduced so that ambiguities in the structure interpretation are eliminated.

The feedback from the wrongly placed atoms in the normal combined synthesis is at a low level. However, as has been shown, some small subtle distortions may be found in some areas of the protein. It is clear, though, that as the model is rebuilt in the areas that are wrong then this distortion decreases. In an attempt to analyse and reduce further the feedback from the misbuilt atoms, use was made of two other Fourier syntheses.

(2) *The combined/omit synthesis*

The inclusion of an atom in a structure-factor calculation is liable to bias any Fourier map computed using those structure factors and phases towards that atom. Therefore, it is common during a refinement to omit small numbers of atoms from the calculation and observe the density which returns for them in a subsequent Fourier map. In their refinements of insulin, Dodson & Vijayan (1980) used an ingenious variant of this method. They divided the asymmetric unit into 'slices' each consisting of several Fourier sections. All the atoms whose centres fell inside such a slice were omitted from the structure factor calculation; a difference map was then calculated for these sections and the fit of the omitted atoms was compared to the density which returned. This process was then repeated with different groups of sections until a complete map of the molecule could be assembled. It was possible only to use the central sections of each slice to avoid the problems of phase contributions from atoms whose centres lay outside the omitted sections but some of whose volumes lay within (these are the 'boundary sections' of Fig. 5). In the combined/omit map the sets of phases obtained from such structure-factor calculations were combined with the isomorphous phases and the resultant combined phase sets were each used to compute Fourier maps for the omitted sections. The terms used in the Fourier syntheses were $|F_{obs}|m \times \exp(i\alpha_{comb/omit})$.

It would be anticipated that, calculated in this manner, the combined/omit Fourier map would contain a very low level of bias from incorrectly placed atoms. The only feedback in this case would be due to correlations between the omitted atoms and those retained in the structure-factor calculation. This synthesis was therefore used to check the degree of bias in the normal combined map. One example of the improvements in the electron density that were verified in this manner was the location of the phenylalanine ring of residue 241. In the isomorphous map (Fig. 6*a*) no electron density could be seen for the side chain of this residue. Density appeared for it at the stage of the phase set COMB3 and the side chain was duly incorporated into the model. The normal combined synthesis with phase set COMB5 (Fig. 6*b*) shows strong density for this side chain. The combined/omit map in the same region is shown in Fig. 6(*c*), the electron density for the phenylalanine ring being quite clear, even though all the atoms in this figure had been omitted from the structure-factor calculation.

A comparison of the combined/omit synthesis and the normal combined synthesis as illustrated in Fig. 6 showed that the apparent resolution of the combined/omit map was lower than that of the normal combined map, thus the carbonyl bulges were more obvious in the latter, and the peak heights higher. However, the close agreement of the two maps showed the reliability of normal synthesis.

(3) *The* 2–1 *synthesis*

The use of a synthesis based on terms $(2F_{obs} - F_{calc}) \times \exp(i\alpha_{calc})$ has been described as an aid to refinement

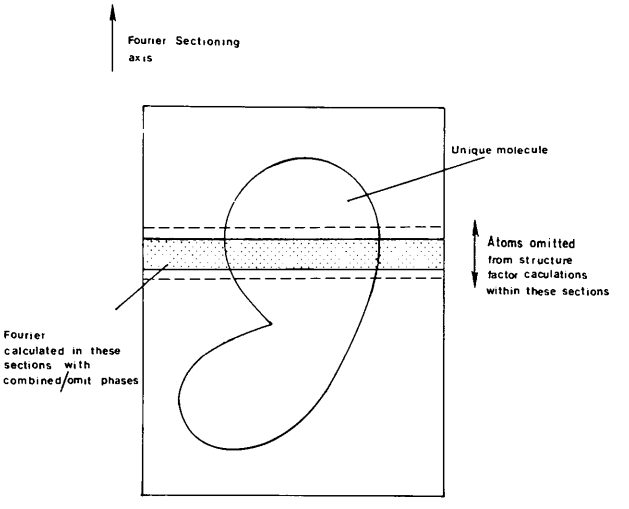


Fig. 5. Schematic diagram designed to illustrate the calculation of a combined/omit phase set.

Fig. 6. The electron density around Phe 241 in three maps; (*a*) Isomorphous 1; (*b*) normal combined synthesis with COMB5 phases; (*c*) the combined/omit map.

(Huber *et al.*, 1974; Main, 1979; Vijayan, 1980). The underlying theory of these maps is that when virtually all of the structure is known and included in α_{calc} then the little that is unknown appears at $\frac{1}{2}$ weight in either $|F_{obs}|\alpha_{calc}$ or $|F_{obs} - F_{calc}|\alpha_{calc}$ maps. Hence the unknown part appears at full weight in $(|F_{obs}| + |F_{obs} - F_{calc}|)\alpha_{calc}$, *i.e.* $|2F_{obs} - F_{calc}|\alpha_{calc}$ maps. This follows from Luzzati (1953). However, if a significant part of the structure is unknown it will appear in $|F_{obs}|\alpha_{calc}$ at less than $\frac{1}{2}$ weight (Luzzati, 1953). The unknown features will then appear at full weight in $[nF_{obs} - (n-1)F_{calc}|\alpha_{calc}$ maps where $n > 2$. Vijayan (1980) has considered the factors which contribute to the best choice of the parameter n and has shown that the best results in terms of peak strengths are obtained when

$$k = \frac{n}{n-1} = 2\langle F_P^2\rangle^{1/2}/\langle F_N^2\rangle^{1/2}, \qquad (7)$$

where N and P represent the total number of atoms and the known number of atoms respectively. In this treatment, however, the errors in the magnitudes of the observed structure factors are not considered and Vijayan (1980) suggests that this necessitates the use of a value of k lower than that given by (7). Thus, when 75% of the scattering matter has been accounted for, $k = 1{\cdot}75$ and therefore, if this value is reduced to 1·5 by the effects of errors, then a map based on coefficients $|3F_{obs} - 2F_{calc}|$ would be appropriate. One disadvantage of such mixed syntheses where n is large is that the error levels in the maps, being the sum of the errors in the $|F_{obs}|\alpha_{calc}$ map plus the errors in the $(n-1)$ difference maps, increase as n increases. It is interesting to note that the factor k is closely related to the factor derived by Luzzati (1953) for the peak heights obtained in difference maps when only a proportion of the structure is known over the range 50–100% known structure.

The question then arises in an analysis of a protein structure at 2·5 Å what proportion of the scattering matter has been satisfactorily accounted for. In assessing this value four factors have to be considered:

(1) the proportion of grossly incorrect structure: for the model PGK7 this represents some 15% of all atoms;

(2) the effects of random positional errors and the inadequacies of the model with respect to the thermal parameters of the atoms: this figure is clearly difficult to estimate but 5% will be considered as a reasonable first approximation;

(3) missing ordered water molecules: following the 1·5 Å refinement of human lysozyme (Artymiuk, 1979), where some 150 water molecules have been located on an enzyme of 1000 atoms, then a figure of 10% might be a good estimate of the effects of missing water, when allowance is made for the generally higher thermal parameters of a proportion of these atoms;

(4) failure of the model to take account of the solvent continuum: again, the contribution of this class of atoms is difficult to estimate but a figure of a further 5% will be used.

The total of these contributions reveals that approximately 35% of the scattering matter might not be represented by a model at an early stage of refinement. Thus it follows that $k = 1{\cdot}6$ and a synthesis with coefficients $|4F_{obs} - 3F_{calc}| \exp(i\alpha_{calc})$ might be appropriate, this synthesis corresponding to a value of k of 1·35. It would also follow from the argument given above that, following an analysis at 2·5 Å, a synthesis with coefficients $|2F_{obs} - F_{calc}|$ would not be expected to be fully effective in suppressing the peaks at wrongly located atoms.

An example of this inability to suppress spurious density is shown in Fig. 7(*a*), which shows the electron density around Phe 291 in a map calculated with coefficients $|2F_{obs} - F_{calc}| \exp(i\alpha_{calc4})$. The normal combined map with COMB5 phases had indicated quite clearly that the phenylalanine ring was misbuilt in the orientation shown in Fig. 7(*a*). However, the synthesis employing calculated phases shown in this figure places quite strong density around the wrongly placed aromatic ring. The correct position for the side chain is indicated by the weak electron density to the right of the present position of the phenylalanine ring. This is unambiguously confirmed in the normal combined map. It is clear that using a $|2F_{obs} - F_{calc}|$ synthesis with calculated phases the error shown here would go undetected. In order to suppress the spurious electron density feature around the incorrectly placed ring a synthesis with coefficients $|5F_{obs} - 4F_{calc}| \exp(i\alpha_{calc4})$ had to be calculated. Although this map gave good density at the true position of the aromatic ring, many distortions of electron density features and false connectivities could be identified in the map due to its inherently high noise level. Thus any interpretation of it requires a degree of hindsight.

However, the possibility exists for producing the same type of mixed synthesis but using the combined phases. Further, it could be expected that the value of n could be reduced using these phases as the whole of the scattering matter in the unit cell is represented by the isomorphous part of the combined phases. Indeed this turned out to be the case as is illustrated in Fig. 7(*b*) which shows the same region of the molecule as in Fig. 7(*a*) but here the electron density map is computed using coefficients $|2F_{obs} - F_{calc}| \exp(i\alpha_{comb5})$. That is, the phases used were those obtained from combination of the calculated phases used to produce Fig. 7(*a*) with the isomorphous phases. The clarity of this map is immediately evident. The electron density at the true side-chain position is very strong and there is no density around the incorrectly located ring. Thus the effect of including the isomorphous phase information is to remove the need to use syntheses where $n > 2$, hence the problems of the noise levels introduced by the use of a greater value of n are avoided.

However, the noise level in a (2–1) map is greater than an equivalent F_{obs} synthesis. The noise decreases the contrast between the protein and the solvent which can lead to confusion in the early stages of an interpretation. Thus the use of this synthesis in the rebuilding was found to be detrimental at the stage where the R factor was in the region of 0·35–0·40. However, below this point the decreased bias in the (2–1) map enabled the maximum amount of information to be extracted at any single stage.

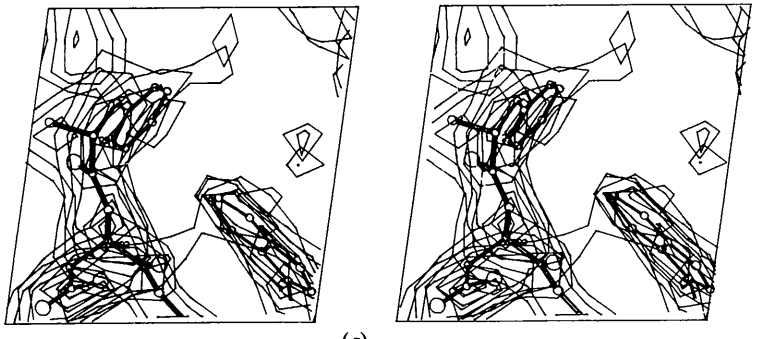
Examination of the complete $(2\text{–}1)\alpha_{combined}$ map showed that it was superior to the equivalent normal combined synthesis. This can be seen from inspection of Fig. 4(*c*) and 4(*d*) which are respectively the normal and (2–1) syntheses with phase set COMB5 corresponding to the model shown in Fig. 4(*c*). In particular, the (2–1) map showed more fine detail than the normal synthesis. Many peaks around the protein surface increased considerably in peak height in the former map. These peaks are found at hydrogen-bonding distances from protein oxygen and nitrogen atoms. When the refinement is extended to higher resolution these may well prove to be well ordered water molecules though such an assignment must remain tentative at this stage.

The general quality of the (2–1) map was further enhanced when the atoms which had been identified as being incorrectly positioned by the use of the above synthesis were removed from the model structure and the remaining atoms used to calculate the phase set COMB6. In comparison with the (2–1) map with phase set COMB5 the one based on these new phases had generally lower levels of background but higher peak heights at the proposed water positions.

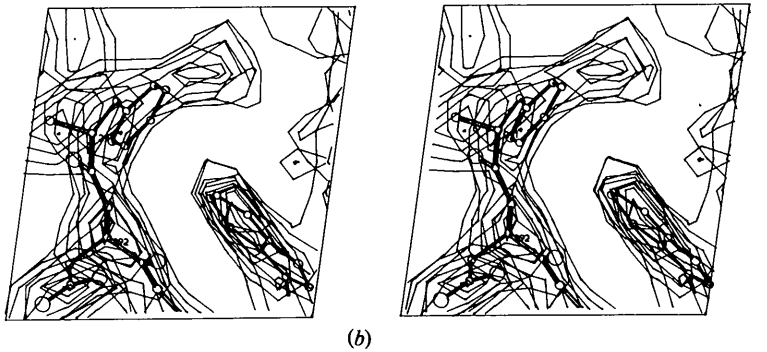
In order to illustrate the overall improvement in the electron density that was obtained, Fig. 8 shows one part of the electron density in the isomorphous map (phase set Isomorphous 2) and in the (2–1) map with COMB6 phases. The improvement in the quality of the density features is evident: the carbonyl bulges in the combined map are very prominent and the side-chain detail greatly enhanced.

Conclusion

The ultimate aim in the study of an enzyme structure is to define accurately the atomic arrangement in the molecule in order to describe its mechanism of action. The conventional procedure of isomorphous replacement has been found to be very limiting in this respect and often higher-resolution data cannot be obtained to enable other techniques to be employed to improve the model of the structure. The experience of the refinement of PGK has shown that, if the initial model of the

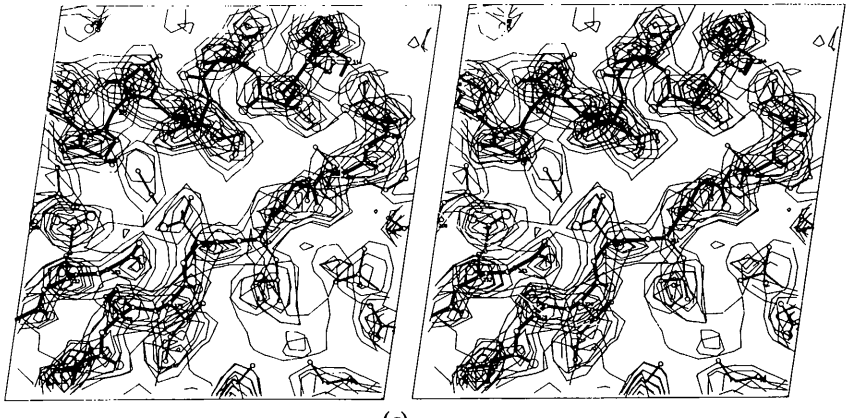


(*a*)



(*b*)

Fig. 7. The use of (2–1) syntheses illustrated in the region of Phe 291: (*a*) (2–1) exp $i\alpha_{calc4}$; (*b*) (2–1) exp $i\alpha_{COMB5}$. The model illustrated is that corresponding to the CALC4 phases. Note the calculated part of the phase set COMB5 is CALC4.



(*a*)

(*b*)

Fig. 8. Stereo diagram illustrating a large area of the molecule in: (*a*) 2–1 map with COMB5 phases; (*b*) Isomorphous map 2.

structure is poor, then at 2·5 Å electron density maps based on calculated phases alone retain a tendency to show spurious positive electron density around incorrectly located atoms, thus hampering the solution of the structure. However, the use of the procedure of phase combination to improve electron density maps has been convincingly demonstrated to have great potential in the field of protein structure determination. The iterative use of the combined phase information is capable of producing very high quality electron density maps far surpassing those obtained by isomorphous replacement alone.

The author wishes to thank Dr Mitchell Lewis for introducing the phase combination program into the laboratory and for discussions on the bias checks and weighting schemes. His thanks are also due to Professor Sir David Phillips for invaluable discussions during the preparation of the manuscript, and to the Medical Research Council for the provision of a project grant to fund the work on PGK.

References

AGARWAL, R. C. & ISAACS, N. W. (1977). *Proc. Natl Acad. Sci. USA*, **74**, 2835–2839.
ARTYMIUK, P. J. (1979). DPhil Thesis, Oxford Univ.
BANKS, R. D., BLAKE, C. C. F., EVANS, P. R., HASER, R., RICE, D. W., HARDY, G. W., MERRETT, M. & PHILLIPS, A. W. (1979). *Nature (London)*, **279**, 773–777.
BLAKE, C. C. F. & EVANS, P. R. (1974). *J. Mol. Biol.* **84**, 585–601.
BLOOMER, A. C., CHAMPNESS, J. N., BRICOGNE, G., STADEN, R. & KLUG, A. (1978). *Nature (London)*, **276**, 362–368.
BLOW, D. M. & CRICK, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.
BLOW, D. M. & ROSSMANN, M. G. (1961). *Acta Cryst.* **14**, 1195–1202.
BRICOGNE, G. (1976). *Acta Cryst.* A**32**, 832–847.
COCHRAN, W. (1951). *Acta Cryst.* **4**, 408–411.
DODSON, E. J., ISAACS, N. W. & ROLLETT, J. S. (1976). *Acta Cryst.* A**32**, 311–315.
DODSON, G. & VIJAYAN, M. (1980). Personal communication.
HARRISON, S. C., OLSON, A. J., SCHUTT, C. E. & WINKLER, F. K. (1978). *Nature (London)*, **276**, 368–373.
HENDRICKSON, W. A. (1971). *Acta Cryst.* B**27**, 1472–1473.
HENDRICKSON, W. A. (1975). *J. Mol. Biol.* **91**, 226–228.
HENDRICKSON, W. A. & KARLE, J. (1973). *J. Biol. Chem.* **248**, 3327–3334.
HENDRICKSON, W. A. & LATTMAN, E. E. (1970). *Acta Cryst.* B**26**, 136–143.
HOPPE, W. & GASSMANN, J. (1968). *Acta Cryst.* B**24**, 97–107.
HUBER, R., KUKLA, D., BODE, W., SCHWAGER, P., HARTELS, K., DEISENHOFER, J. & STEIGEMANN, W. (1974). *J. Mol. Biol.* **89**, 73–101.
KARLE, J. & HAUPTMAN, H. (1956). *Acta Cryst.* **9**, 635–651.
KONNERT, J. H. (1976). *Acta Cryst.* A**32**, 614–617.
LUZZATI, V. (1953). *Acta Cryst.* **6**, 142–152.
MAIN, P. (1979). *Acta Cryst.* A**35**, 779–785.
RANGO, C. DE, MAUGUEN, Y. & TSOUCARIS, G. (1975). *Acta Cryst.* A**31**, 227–233.
RICE, D. W. (1979). DPhil Thesis, Oxford Univ.
ROSSMANN, M. G. & BLOW, D. M. (1961). *Acta Cryst.* **14**, 631–647.
SAYRE, D. (1974). *Acta Cryst.* A**30**, 180–184.
SIM, G. A. (1959). *Acta Cryst.* **12**, 813–815.
SIM, G. A. (1960). *Acta Cryst.* **13**, 511–512.
VIJAYAN, M. (1980). *Acta Cryst.* A**36**, 295–298.

Acta Cryst. (1981). A**37**, 500–506

Representation of Bicrystal Invariant Translations

BY A. BROKMAN*

Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, *USA*

(*Received* 8 *September* 1980; *accepted* 15 *January* 1981)

Abstract

All complete pattern shift lattice (DSC-lattice) translations of one lattice with respect to the second lattice of a bicrystal are described as a group. It is shown that the matrix representation of this group can be used to solve topological problems connected with secondary grain boundary dislocations (SGBD's) such as finding the step in the boundary associated with the SGBD. We have formulated this problem by establishing the 'step vector' **S** associated with the Burgers vector **b** of the SGBD in a cubic bicrystal. The problem is then solved in $2D$, and the way to generalize to $3D$ is indicated.

* Also at Bergman School for Applied Science, Hebrew University of Jerusalem, Israel.

0567-7394/81/040500-07$01.00